



HAL
open science

Annotation automatisée des métadonnées structurelles dans les partitions musicales : cas des modulations et des cadences pour la forme sonate

Laurent Feisthauer

► To cite this version:

Laurent Feisthauer. Annotation automatisée des métadonnées structurelles dans les partitions musicales : cas des modulations et des cadences pour la forme sonate. Algorithme et structure de données [cs.DS]. Université de Lille, 2021. Français. NNT : 2021LILUI032 . tel-03471256

HAL Id: tel-03471256

<https://theses.hal.science/tel-03471256>

Submitted on 8 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse

Préparée pour l'obtention
du Doctorat de l'Université de Lille,
spécialité Informatique,
par

Laurent Feisthauer

Annotation automatisée des métadonnées structurelles dans les partitions musicales

Cas des modulations et des cadences pour la forme sonate

Thèse soutenue le 18 mai 2021, devant le jury constitué de

Marc Rigaudière	Professeur, Univ. Reims Champagne-Ardennes	<i>rapporteur</i>
Philippe Rigaux	Professeur, CNAM	<i>rapporteur</i>
David R. W. Sears	Assistant Professor, Texas Tech University	<i>examineur</i>
Marc Tommasi	Professeur, Université de Lille	<i>examineur</i>
Anja Volk	Assistant Professor, Utrecht University	<i>examinatrice</i>
Louis Bigo	Maître de conférences, Univ. de Lille	<i>co-encadrant</i>
Mathieu Giraud	Directeur de recherches, CNRS	<i>co-directeur</i>
Florence Levé	Maître de conférences, UPJV, Amiens	<i>co-directrice</i>

Remerciements

Je souhaite remercier l'ensemble du jury pour leur disponibilité et leur bienveillance à mon égard et à celui de mon travail. Merci au professeur Marc Tommasi, président du jury, pour la conduite de la soutenance ainsi que de m'avoir donné l'opportunité d'enseigner pendant 2 ans au sein de l'UFR MIME. Merci aux professeurs Marc Rigaudière et Philippe Rigaux, rapporteurs de cette thèse, pour le temps passé à la lecture de cette thèse et à la production de leur rapport dont la lecture m'a rendu particulièrement heureux. Merci également de leur disponibilité avant, pendant et même après la soutenance. *Thank you to the doctor David R. W. Sears and to the doctor Anja Volk, to honor me by their presence as members of my thesis defense committee, knowing that it would be defended in a language that they are not that familiar with. You have my deepest gratitude.*

Merci à mes encadrants, Louis Bigo, Mathieu Giraud et Florence Levé, pour m'avoir offert l'opportunité d'une vie et pour m'avoir accompagné durant cette aventure et m'avoir poussé dans mes retranchements pour fournir le meilleur travail possible. Ces presque 4 ans ont été remplis d'expériences inoubliables, complexes, parfois pénibles mais toujours enrichissantes et vous ne m'avez jamais laissé tomber et avez toujours pris le temps de me soutenir et de répondre à mes questions et de lever mes doutes, même quand vous étiez vous même forts occupés et ce quels qu'en soient les motifs. J'espère que nous aurons à nouveau l'occasion de travailler ensemble dans le futur.

Merci à Emmanuel, Nicolas, Richard, Pierre, Jules, Gianluca, Louis (les 3), David, Quentin, Benoît et Viet-Toan de l'équipe **Algomus** pour leur accueil et leur gentillesse au quotidien. Je garde d'excellents souvenirs de nos discussions et des journées au vert. J'aurais aimé avoir l'occasion de plus vous connaître mais le COVID en a décidé autrement. Je vous souhaite une excellente continuation dans vos projets.

Merci à toutes les personnes du laboratoire **CRISTAL** avec qui j'ai pu échanger pour leur disponibilité et en particulier à ma collègue doctorante Emilie pour les pauses cafés et les discussions variées et toujours passionnantes.

*I would also like to thank everyone on the **DDMAL** team that I visited on summer 2019, thanks to Pr. Ichiro Fujinaga. I hope that the CTS team will publish a lot of delightful papers in the future. I want to give a special thanks to my local key buddy Néstor. I'm very proud of our paper and our collaboration.*

J'aimerais enfin remercier mes amis et ma famille pour leur soutien tout au long de cette aventure. Je remercie en particulier ma compagne Luce pour sa patience durant les moments difficiles et son soutien inconditionnel.

À la mémoire de Fernand Feisthauer

À mon fils Oscar

Cette thèse vous est dédiée

Table des matières

1	Introduction	7
2	Tonalité, cadence et forme, des concepts musicaux haut-niveau pour l'analyse des partitions	11
2.1	Un système pour organiser les notes : le système tonal	12
2.1.1	Émergence	12
2.1.2	Définition et éléments caractéristiques	13
2.1.3	Propriétés du système tonal	13
2.2	Tonalité, parcours tonal, modulations et tonicisations	14
2.3	Un processus de terminaison des phrases musicales : la cadence	17
2.3.1	Principes fondamentaux	17
2.3.2	Types de cadences	18
2.4	Structurer la musique de façon dialectique : la forme sonate	20
2.4.1	La forme sonate, une organisation tonale et thématique	21
2.4.2	Théories récentes sur la forme sonate	21
2.4.3	Préparation et articulation de la césure médiane	24
2.5	Bilan	26
3	État de l'art en musicologie numérique sur l'estimation de la tonalité et sur la détection des cadences	27
3.1	Estimation automatisée de la tonalité	28
3.1.1	L'algorithme de Longuet-Higgins et Steedman	30
3.1.2	Carol L. Krumhansl et les <i>profils tonaux</i> (<i>pitch profiles</i>)	30
3.1.3	Monopole des profils tonaux pour l'estimation de tonalité	32
3.1.4	Autres méthodes pour l'estimation de tonalités	35
3.2	Perception, modélisation et détection automatisée des cadences	37
3.2.1	Cadences et perception de fermeture	38
3.2.2	Modélisation des cadences pour la prédiction automatisée	39
3.3	Bilan	40

4	Encodage et annotation de corpus pour la musicologie computationnelle	41
4.1	Corpus annotés de cadences préexistants	43
4.2	Corpus des quatuors à cordes de Mendelssohn : la qualité des données avant tout!	43
4.2.1	Contexte	43
4.2.2	Comparaison des différents stades de corrections	47
4.2.3	Annotations	48
4.3	Corpus des quatuors à cordes de W. A. Mozart : des annotations pour l'analyse musicale	48
4.3.1	Contexte	48
4.3.2	W. A. Mozart et les quatuors à cordes	49
4.3.3	Annotations	50
4.3.4	Utilisation du corpus	55
4.4	Modulations et tonicisations	55
4.4.1	Contexte	55
4.4.2	Références théoriques sélectionnées	55
4.4.3	Encodage et annotations	56
4.4.4	Statistiques	57
4.5	Bilan	59
5	Modélisation et estimation de la tonalité et des modulations	61
5.1	Trois critères pour une modulation	62
5.2	Modélisation des trois critères	64
5.2.1	Ancrage dans la tonalité : évaluer la proximité d'une progression harmonique caractéristique, $c_{V \rightarrow I}(b, k)$	65
5.2.2	Compatibilité de la tonalité avec le texte musical : $d_{diat}(\mathcal{S}, \mathcal{S}')$	66
5.2.3	Proximité entre les tonalités : garantir une certaine stabilité tonale, $d_W(k, k')$	69
5.3	Combinaison des trois critères pour estimer le parcours tonal et les modulations	70
5.4	Corpus et méthodologie d'évaluation	71
5.5	Application (Mozart)	73
5.5.1	Discussion sur les mesures	73
5.5.2	Détection des progressions $V \rightarrow I$	73
5.5.3	Évaluation de la mesure d'ancrage de la tonalité sur K157.3	74
5.5.4	Compatibilité entre gammes	75
5.5.5	Estimation des tonalités et détection des modulations	76
5.5.6	Discussion	77
5.6	Application (modulations et tonicisations)	78
5.6.1	Modèles évalués	79
5.6.2	Discussion	80
5.7	Bilan	82

6	Modélisation et détection des cadences	83
6.1	Un problème de classification	84
6.2	Descripteurs pour les cadences	86
6.2.1	Descripteurs du potentiel accord d'arrivée de la cadence.	86
6.2.2	Descripteurs rythmiques à proximité du point d'arrivée Z	90
6.2.3	Descripteurs directement avant le point d'arrivée cadentiel et aux alentours.	90
6.2.4	Descripteurs sur le passé plus lointain de la cadence	91
6.2.5	Statistiques sur les descripteurs	91
6.3	Détection de PAC et de HC chez Bach et Haydn	93
6.3.1	Corpus et implémentations	93
6.3.2	Processus d'apprentissage	94
6.3.3	Discussion	94
6.3.4	Bilan	96
6.4	Descripteurs pour une cadence particulière : la césure médiane	96
6.4.1	Des descripteurs pour la césure médiane	97
6.4.2	Évaluation des descripteurs sur le corpus des quatuors de Mozart	98
6.5	Détection des césures médianes chez Mozart	99
6.5.1	Processus d'apprentissage	99
6.5.2	Détection des césures médianes	99
6.6	Bilan	100
7	Conclusion	103
	Bibliographie	105
	Résumé	116
	<i>Abstract</i>	116

Chapitre 1

Introduction

Cette thèse s’inscrit dans le domaine de *l’informatique musicale* et plus particulièrement *l’analyse musicale computationnelle*. S’il peut sembler incongru de prime abord d’associer la musique, art de l’ineffable par excellence [69], et l’informatique, qui ne tolère aucune expression imprécise, de nombreuses personnes ont proposé de les lier et ce dès les prémices de l’informatique. En 1843, A. Lovelace, première personne à concevoir un programme informatique émet l’hypothèse que “la machine pourrait composer de manière scientifique et élaborée des morceaux de musique de n’importe quelle longueur ou degré de complexité” pour peu qu’on puisse exprimer de façon analytique les relations fondamentales entre les notes dans la science de l’harmonie et de la composition musicale [81]. L. Hiller et L. Isaacson valident cette hypothèse en 1951 en concevant le premier quatuor à cordes (et première œuvre musicale) entièrement composé par l’ordinateur, *Illiad Suite* [66].

L’informatique musicale manipule différents objets tels que des fichiers sons (signaux audio), des images, des vidéos, des partitions, des métadonnées ou des informations en temps réel [8]. Cette thèse s’inscrit dans l’étude de *partitions musicales* représentées *symboliquement* par la notation musicale et encodées informatiquement par des formats tels que MIDI, krn, MusicXML ou MEI.

L’analyse musicale consiste à étudier une œuvre pour en définir la forme, la tonalité, la structure, le rythme, l’harmonie, l’orchestration, la thématique, la mélodie, la dynamique, etc. de façon à en comprendre sa genèse, son sens, son but mais aussi d’expliquer le geste du compositeur et guider l’interprétation [13].

Ces éléments sont autant d’annotations (ou *métadonnées*) musicales qui sont déduites (ou *extraites*) des notes, des rythmes et des indications sur la partition – quand il y a une partition, ce qui est le cas pour les musiques tonales étudiées ici. L’extraction de données est une opération très courante dans le domaine informatique et c’est pourquoi l’analyse musicale semble possible via l’ordinateur. Les théoriciens de la musique ont toujours cherché à avoir une vue haut-niveau de la musique, que ce soit sur l’harmonie, la forme, ou d’autres concepts. Le domaine de l’analyse musicale computationnelle (*Computational Music Analysis, CMA*) tente, avec l’aide de l’ordinateur, de modéliser les partitions musicales ainsi que les différents concepts qui gouvernent la musique [85]. Dès les années 1960, des théoriciens de la musique utilisent des approches calculatoires pour proposer des notions d’analyse, comme, par exemple, la *Set Theory* développée par M. Babbitt, G Perle, A. Forte et E. Carter [50].

Une étape importante pour l'analyse musicale computationnelle est la parution en 1983 du livre *A Generative Theory of Tonal Music* (GTTM) écrit conjointement par le musicologue F. Lerdahl et le linguiste R. Jackendoff [79], influencés par l'analyse schenkérienne du début du XX^e siècle [120]. Les auteurs y proposent 56 règles permettant non seulement la segmentation de la musique tonale de deux façons (suivant la structure et suivant l'alternance de temps forts et de temps faibles) mais aussi la réduction d'une partition musicale selon ces deux segmentations. La segmentation puis la réduction conduisent à la génération d'un arbre qui représente la structure de l'œuvre. Un exemple de ces segmentations ainsi que de la réduction est observable à la figure 1.1. Cet ensemble de règles forme une grammaire générative de la musique. Le nombre de règles est conséquent et celles-ci sont souvent imprécises et ambiguës. Cependant, la représentation de la musique sous forme d'arbre, qui est une structure de données bien connue des informaticiens, a inspiré certains travaux d'informatique musicale notamment concernant la modélisation des progressions harmoniques [113] ou du rythme [52]. La GTTM est ainsi une modélisation à haut-niveau de la musique. Certains auteurs ont modélisé et implémenté des aspects de la GTTM sur de courts extraits [61], mais l'analyse computationnelle complète haut-niveau d'une œuvre musicale n'est pas du tout évidente, à cause de la complexité algorithmique de la méthode et de l'ambiguïté des règles.

Cette thèse propose de nouvelles approches basées sur l'apprentissage machine et la programmation dynamique pour modéliser et analyser deux concepts spécifiques à la musique, la *tonalité* et les *cadences*. En particulier, elle s'inscrit dans le défi de parvenir à une analyse complète d'une œuvre musicale en forme sonate qui est l'un des sujets de recherche de l'équipe *Algomus* dans laquelle j'ai mené cette thèse.

Le chapitre 2 est consacré à la définition de notions musicologiques et en particulier, la tonalité et les modulations, les cadences et la forme sonate. Ces définitions sont établies en s'appuyant sur plusieurs ouvrages musicologiques de référence.

Le chapitre 3 présente la littérature existante dans le domaine de la musicologie computationnelle pour estimer la tonalité globale ainsi que la tonalité locale en tous points d'une œuvre musicale, avec un focus sur le *profil tonal*. Nous nous penchons également sur la perception de la sensation de fermeture engendrée par les cadences et sur les stratégies mises en place pour détecter les points d'arrivée cadentiels.

Les contributions propres de la thèse sont présentées dans les chapitres 4, 5 et 6.

Cette thèse a contribué à la création de trois corpus distincts qui sont présentés dans le chapitre 4. Le premier corpus concerne l'encodage, c'est-à-dire la transcription sous format numérique, des partitions des six *quatuors à cordes* de F. Mendelssohn ainsi que de quatre *pièces isolées pour quatuor à cordes* du même compositeur, avec un travail sur l'évaluation de la qualité de ces encodages. Le deuxième corpus consiste en l'annotation de mouvements issus des *quatuors à cordes* de W. A. Mozart. Les annotations indiquent les tonalités et modulations, les points d'arrivée cadentiels et la forme du mouvement. Le troisième corpus est un corpus de *modulations* constitué à partir d'extraits provenant de cinq livres de théorie musicale. Ce corpus est donc spécialisé dans l'estimation de la tonalité locale et l'identification des points de modulation.

Le chapitre 5 présente un algorithme d'estimation des tonalités en tout point de la partition. Cet algorithme estime à chaque temps de la partition trois distances correspondant à trois signaux musicaux permettant d'identifier les points de modulation, à savoir *l'ancrage dans la*

FIGURE 1.1 – Figure extraite de *A Generative Theory of Tonal Music* [79, p. 144]. Les auteurs proposent une réduction temporelle du choral “O Haupt voll Blut und Wunden” issue de la Passion selon Saint Matthieu de Bach. Au dessus de la partition, l’arbre représente la structure hiérarchique de l’œuvre. Il est construit selon les règles de structure et de métrique. Les portées en dessous de la partition correspondent aux différents niveaux de réduction.

tonalité, la compatibilité des notes et la proximité entre les tonalités. Le parcours tonal s’estime par recherche du chemin optimal par programmation dynamique, une fois les distances estimées pour tous les temps et pour toutes les tonalités. Une première expérience utilisant cette approche est menée sur le corpus des quatuors de Mozart. Elle permet d’estimer correctement la tonalité sur près de 85% des temps du corpus. Une deuxième expérience est menée sur le corpus spécialisé, expérience dans laquelle le modèle est comparé à deux autres modèles récents d’estimation de tonalité locale.

Le chapitre 6 propose dans une première partie un algorithme de détection des cadences par l’extraction de descripteurs de haut-niveau caractéristiques de la présence d’un point d’arrivée cadentiel sur la partition musicale. La significativité de chacun de ces descripteurs est étudiée

puis ceux-ci servent à entraîner un algorithme d'apprentissage qui classe chaque temps de la partition comme un point d'arrivée cadentiel ou non. La détection de cadences parfaites sur les fugues de Bach et les quatuors de Haydn est satisfaisante (F_1 -mesure de 0.80 pour Bach et de 0.69 pour Haydn mais plus de 80% de précision), mais peut être améliorée pour les demi-cadences (F_1 -mesure de 0.29 sur Haydn due à un nombre très élevé de faux positifs). Une deuxième partie s'attarde sur la détection d'une cadence particulière caractéristique de la forme sonate : la *césure médiane*. Là encore, une extraction de descripteurs de haut-niveau précède un algorithme d'apprentissage machine. Les descripteurs sont encore plus abstraits – la césure médiane étant de plusieurs types différents, elle est caractérisée par certains signaux qui peuvent s'étaler sur plusieurs mesures. Malgré un ensemble d'entraînement de taille très modeste, le modèle localise correctement la césure médiane pour la moitié du corpus.

La conclusion (chapitre 7) fait un bilan de ces travaux et propose des pistes pour approfondir et élargir le travail de cette thèse, que ce soit sur les formes sonates ou sur d'autres expressions musicales.

Chapitre 2

Tonalité, cadence et forme, des concepts musicaux haut-niveau pour l'analyse des partitions

Contenu

2.1	Un système pour organiser les notes : le système tonal	12
2.1.1	Émergence	12
2.1.2	Définition et éléments caractéristiques	13
2.1.3	Propriétés du système tonal	13
2.2	Tonalité, parcours tonal, modulations et tonicisations	14
2.3	Un processus de terminaison des phrases musicales : la cadence	17
2.3.1	Principes fondamentaux	17
2.3.2	Types de cadences	18
2.4	Structurer la musique de façon dialectique : la forme sonate	20
2.4.1	La forme sonate, une organisation tonale et thématique	21
2.4.2	Théories récentes sur la forme sonate	21
2.4.3	Préparation et articulation de la césure médiane	24
2.5	Bilan	26

Dans ce chapitre, nous présentons certaines notions musicales : la tonalité, les cadences et la forme sonate. Comme ces notions sont au cœur du reste de ce manuscrit, il nous semble important de les introduire au lecteur. Ce chapitre n'a pas la prétention de présenter en détail ces notions musicales complexes mais plutôt de préciser quelles propriétés nous avons retenues, pour pouvoir en proposer une modélisation informatique par la suite.

Une notion est centrale à cette thèse, la *tonalité*. Ce mot est ambigu en français puisqu'il désigne deux choses [110, p. 125]. La *tonalité comme système*, en anglais *tonality*, décrite dans la section 2.1, désigne le langage qui organise la musique à partir du XVIII^e siècle. Nous nous y référerons tout au long de cette thèse par l'expression *système tonal*. La *tonalité* qui peut-être *globale ou locale*, en anglais *key*, décrite dans la section 2.2, est la concaténation du nom de la tonique avec le mode utilisé (comme *Do Majeur*) pour définir la gamme utilisée et les fonctions de chaque accord. Nous réservons l'expression *tonalité* pour cette définition à l'exception des usages faits dans la section 2.1 et ce afin d'être fidèle aux mots utilisés par les musicologues cités.

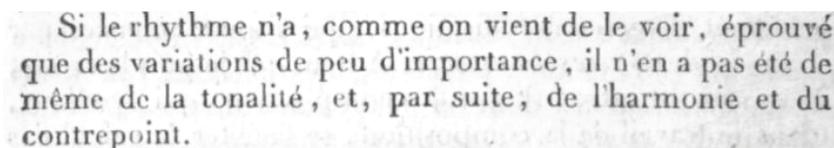
L'œuvre musicale s'organise sous forme de phrases. La fin de ces phrases s'articule de manière codifiée par la présence de progressions harmoniques caractéristiques. On appelle l'articulation de ces fins de phrases, les *cadences*. Nous les présentons dans la section 2.3.

Ces phrases structurent l'œuvre musicale en différentes sections qui en dessinent la forme. Nous nous intéressons en particulier à la *forme sonate* dans la section 2.4. Elle est indissociable du système tonal par la dissonance à grande échelle qu'elle implique [115].

2.1 Un système pour organiser les notes : le système tonal

2.1.1 Émergence

Le terme "tonalité" apparaît pour la première fois en 1810, dans le sommaire de l'histoire de la musique préfaçant le *Dictionnaire historique des musiciens, artistes et amateurs, morts ou vivans* écrit par A. Choron et F.-J.-M. Fayolle [29] (Figure 2.1).



Si le rythme n'a, comme on vient de le voir, éprouvé que des variations de peu d'importance, il n'en a pas été de même de la tonalité, et, par suite, de l'harmonie et du contrepoint.

FIGURE 2.1 – Le mot "tonalité", la première fois qu'il est employé, en 1810, désigne la concaténation de la tonique et du mode. Extrait de [29].

Cependant, le concept de tonalité (ou plus largement de système tonal) apparaît bien avant. Il commence à prendre forme autour des années 1600 et coïncide avec le début de l'ère baroque. Le système tonal est bien établi à l'époque classique (1730-1830) et sera enrichi jusqu'à atteindre sa complexité maximale pendant l'ère romantique (1830-1910) jusqu'à son abandon par les compositeurs de musique dite "savante" au XX^e siècle.

L'un des écrits fondamentaux pour la définition de la tonalité est le *Traité de l'harmonie réduite à ses principes naturels* [105] écrit par le compositeur, maître de chapelle et théoricien de la musique J.-P. Rameau en 1722. L'auteur énonce dans cet ouvrage le principe de l'équivalence des octaves, la notion de basse fondamentale et de renversement des accords, la primauté de

l'accord parfait majeur et de l'accord parfait mineur. Il pose ainsi les bases de l'harmonie classique et du système tonal d'une manière qui n'est plus empirique.

2.1.2 Définition et éléments caractéristiques

À l'entrée "tonalité" du *dictionnaire de la musique* sous la direction de M. Vignal [134], celle-ci est définie comme¹ :

"le caractère propre à toute musique fondée, dans le maniement des hauteurs, sur le principe d'une *hiérarchie* entre les différents degrés de hauteur, donnant à certains d'entre eux, et surtout à l'un, la *tonique* (ou premier degré, I), le statut privilégié de *notes attractives* vers lesquelles tendent les autres degrés, et sur lesquels on se repose."

La tonalité est donc un système musical au sens de C. Dahlhaus [37]; "une échelle musicale, c'est-à-dire l'ensemble des notes qu'utilise un type d'écriture donné, et une conception de la musique, qui transforme ce matériau sonore en un complexe de relations musicales."

Le système tonal s'organise en plusieurs éléments caractéristiques complexes. Le plus immédiat est l'échelle, appelée *gamme*, polarisée autour d'une note appelée *tonique*. Cette échelle consiste en une succession d'intervalles choisie selon le mode considéré, *majeur* ou *mineur*. Cette échelle est toujours *heptatonique* et *diatonique*, elle contient sept notes dont les noms sont tous différents.

Ces notes peuvent se succéder dans la temporalité musicale², mais peuvent aussi être jouées simultanément, ce qu'on appelle *accord*. On associe à chaque degré de la gamme un accord d'au moins trois sons construit comme une superposition de tierces. On ajoute à cet ensemble d'accords certains accords particuliers tels que l'accord de sixte sur degré napolitain ou encore l'accord de sixte augmentée qui observent d'autres règles de construction.

Les accords peuvent se succéder dans la temporalité, créant une *progression harmonique*, l'*harmonie* étant la science des enchaînements d'accords. Ceux-ci s'inscrivent dans un réseau de relations définies par leur degré dans la gamme. Les accords des premiers, quatrièmes et cinquième degrés sont les plus prégnants, autant par leur nombre que par leur importance [3]. On leur attribue des noms spécifiant leur fonction : la *tonique* pour le premier degré, la *dominante* pour le cinquième degré et la *sous-dominante* pour le quatrième degré. On rapporte les autres degrés à ces trois fonctions principales.

2.1.3 Propriétés du système tonal

Plusieurs théoriciens observent différentes propriétés qui permettent de saisir le système tonal dans son ensemble.

C. Accaoui et M. Joos, dans les *Éléments d'esthétique musicale* [4], énoncent trois propriétés pour le système tonal, qu'ils désignent par "tonalité" ci-après.

La tonalité est *rationnelle* parce que son fonctionnement est celui d'un système. Ce système est *englobant* car tous les sons potentiels existent dans ce système, *centré* car tous les sons s'or-

1. Il s'agit de la première définition du mot tonalité dans l'ouvrage. Deux autres définitions sont données. L'une est historique (limitation à deux modes, etc.). L'autre concerne la tonalité dans laquelle est écrit un morceau et implique un aspect harmonique. Cette définition est développée dans la section 2.2.

2. Ce qu'on appelle *mélodie* mais qui n'est bien entendu pas restreint à la notion de tonalité. La mélodie est d'ailleurs bien plus qu'une succession de notes, on lui attache aussi des caractéristiques plastiques.

ganisent par rapport à un accord central de tonique du ton principal et *homogène* car tout son du système s’englobe dans l’englobement et le centrage du système.

La tonalité est *expressive*. Non seulement elle fait ressortir le schéma rythmique de tension/détente et élan/retombée par les progressions harmoniques, mais en donnant une primauté à l’harmonie, elle met en avant la voix du dessus, celle de l’individu qui chante.

Enfin, la tonalité caractérise un *temps circulaire et directionnel*. Un morceau commence dans une tonalité principale, puis explore de nouvelles tonalités avant de revenir à la tonalité principale. Il est donc circulaire. Mais la notion de tensions à résoudre le rend également directionnel. Une sensible ne peut qu’aller vers la tonique, l’accord de dominante se résout sur l’accord de tonique.

D. Tymoczko dans *The geometry of music* identifie cinq caractéristiques du sentiment de système tonal [132] :

- des *mouvements mélodiques conjoints*, on privilégie les petits intervalles à l’intérieur des voix.
- une *consonance acoustique*, on préfère les harmonies consonnantes (comme celle de tierce) et celles-ci marquent les points de stabilité musicale.
- une *cohérence harmonique*, les harmonies d’un passage musical sont structurellement similaires les unes par rapport aux autres.
- une *macro-harmonie limitée*, on n’utilise qu’un certain ensemble de notes sur une période limitée dans le temps.
- une *centricité* autour de la tonique qui est la note préminente d’une tonalité donnée.

Si ces théoriciens s’accordent sur certaines de ces notions telles que la centricité et son homogénéité (rapprochée avec la cohérence harmonique), ils s’opposent sur ce qui guide le temps musical tonal. Pour D. Tymoczko, c’est la conduite de voix qui engendre le discours, étant elle-même engendrée par la macro-harmonie sous-entendue par la tonalité ainsi que les règles de consonance acoustique. Pour C. Accaoui et M. Joos, le discours musical est guidé par la mélodie et les accords-degrés qui la supportent. L’expressivité de la mélodie se retrouve également dans le parcours tonal qui a pour but à court terme d’explorer de nouvelles tonalités avant de revenir à long terme à son point de départ, c’est l’harmonie qui guide le discours.

Cette discordance rend la modélisation informatique du système tonal difficile. Cependant, certaines de ses propriétés sont modélisables informatiquement afin d’en étudier la validité. F. C. Moss et al. observent statistiquement certaines de ces caractéristiques dans les quatuors à cordes de L. V. Beethoven [90], en particulier la centricité, la référentialité (certaines progressions harmoniques sont privilégiées), l’unidirectionnalité des progressions harmoniques et la hiérarchisation des degrés de la gamme.

2.2 Tonalité, parcours tonal, modulations et tonicisations

Nous discutons maintenant de la tonalité comme le nom donnant la “clé” définissant les relations hiérarchiques entre les hauteurs de notes et les progressions harmoniques privilégiées.

Une œuvre de musique tonale commence et s’achève dans une unique tonalité appelée *tonalité principale*³. Cette tonalité caractérise le morceau et peut même apparaître dans le nom

3. Il existe certaines œuvres qui ne se terminent pas dans la tonalité initiale, je considère celles-ci comme des exceptions à cette règle. On peut toujours expliquer ce phénomène par l’analyse du texte musical.

The image shows a musical score for the beginning of the 4th movement of the 15th string quartet by W.A. Mozart, measures 1 to 16. The score is in 8/8 time and features a modulation from D minor to F major. Three tonicizations are circled in black: measures 3-7 (D major/D minor), measure 6 (D major), and measure 15 (F major). A blue highlight covers measures 10-16, labeled 'm10'.

FIGURE 2.2 – Début du 4^e mouvement du 15^e quatuor à cordes en Ré mineur K. 421 de W. A. Mozart, mesures 1 à 16. Deux tonalités différentes se succèdent au cours de cet extrait. Le morceau commence en Ré mineur, qui est donc la tonalité principale de ce mouvement (partie saumon), puis module en Fa Majeur à la mesure 10 (partie mauve). Trois tonicisations sont entourées en noir. Les deux premières, mesures 3 et 6-7, empruntent à Sol Majeur ou Sol mineur. La troisième, mesure 15, use de l'accord de dominante de dominante en Fa Majeur et tonicise donc en Do Majeur.

de celui-ci, comme dans la *Sonate en Si mineur* de F. Liszt. Comme elle concerne le morceau dans sa globalité, on peut également l'appeler *tonalité globale*.

Le *parcours tonal* présente les différentes tonalités empruntées au cours du morceau. On peut le modéliser sous la forme suivante :

$$\textit{tonalité}_{\textit{principale}} \rightarrow \textit{tonalité}_1 \rightarrow \dots \rightarrow \textit{tonalité}_n \rightarrow \textit{tonalité}_{\textit{principale}}$$

Tonalités locales et cycle des quintes. On appelle chacune des tonalités empruntées par l'œuvre musicale une *tonalité locale*. Ainsi, on peut associer une tonalité locale en chaque point de la partition, qui ne sera pas forcément la tonalité principale.

Sur la figure 2.2, la tonalité principale ou globale est celle du début du mouvement, c'est-à-dire *Ré mineur*. Par contre, on a deux tonalités différentes dans l'extrait. Les mesures 1 à 9 sont en *Ré mineur* et les mesures 10 et suivantes sont en *Fa Majeur*. Ainsi la tonalité locale à tout moment de la mesure 11 ou de la mesure 12 sera *Fa Majeur*.

L'acte de passer d'une tonalité à une autre s'appelle la *modulation* qui, contrairement à ce que son nom indique, n'est pas juste un changement de mode mais aussi de tonique⁴. Comme établir une nouvelle tonalité locale, c'est aussi changer les relations et les fonctions de chacun des degrés, la modulation doit être préparée et la tonalité d'arrivée reste la tonalité locale pendant quelques mesures. En général, la modulation a lieu entre deux tonalités ayant une relation privilégiée.

Plusieurs systèmes définissent la proximité entre tonalités, comme le *cycle des quintes* (figure 2.3). Dans ce cycle, les tonalités ayant une relation privilégiée sont adjacentes. En assimilant les notes enharmoniques comme équivalentes (hypothèse de tempérament égal; Mi \flat = Ré \sharp) et en partant d'une note choisie, l'enchaînement de douze quintes dans un même sens (ascendant ou descendant) permet de passer par toutes les notes de la gamme chromatique puis de revenir à la note de départ. On représente cette périodicité par un cercle. Ce cycle a été théorisé tout d'abord par le compositeur et théoricien ukrainien N. Diletski dans son *Idea grammatikii musikiyskoy* [40] puis remanié par le compositeur et théoricien allemand J. D. Heinichen, qui lui donna la forme que l'on connaît aujourd'hui dans son traité de 1711, *Der General-Bass in der Composition* [62] en associant aux notes les tonalités s'y rapportant (majeures et mineures). Le cycle des quintes et un autre système de mise en relation des tonalités sont discutés ultérieurement dans la thèse (section 5.2.3).

Tonicisations. Il est possible d'emprunter très brièvement une tonalité adjacente dans le cycle des quintes afin de mettre en évidence un degré particulier de la tonalité locale. Par exemple, sur la figure 2.2 à la dernière croche de la mesure 6, un accord de septième de dominante de *Sol Majeur* est joué sous son premier renversement (les notes sont du grave vers l'aigu *Fa \sharp , La, Do \sharp , Ré*). Or cet accord n'existe pas sur un degré de *Ré mineur* (car *Fa \sharp* ainsi que *Do \sharp* ne sont pas dans la gamme harmonique de *Ré mineur*). On emprunte ainsi en *Sol Majeur* (ou *Sol mineur*) sur la dernière croche de la mesure 6, tonalité dont on entend le premier degré I sur le premier temps de la mesure 7. On pourrait penser qu'on module en *Sol Majeur* mais il n'en est rien. On revient tout de suite en *Ré mineur* sans ambiguïté puisqu'on a une cadence parfaite en *Ré mineur* dont le point d'arrivée est sur le premier temps de la mesure 8.

4. C'est vrai dans le contexte de la tonalité. J. P. Rameau considère la modulation comme un changement de mode uniquement [105].

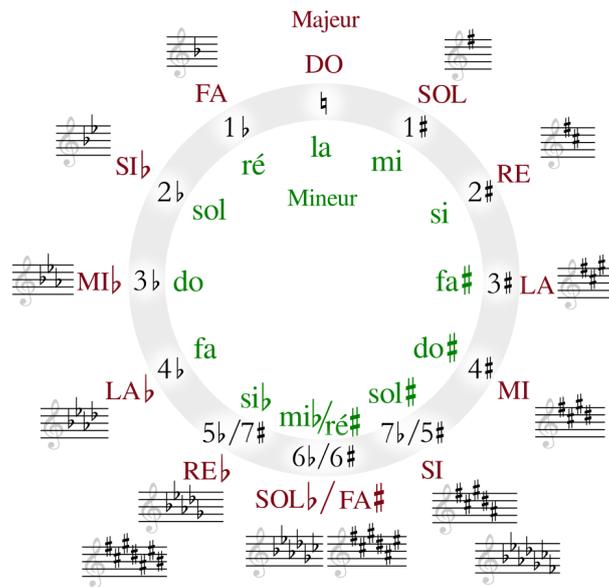


FIGURE 2.3 – Cycle des quintes complet issu de wikipedia : https://commons.wikimedia.org/wiki/File:Circle_of_fifths_deluxe_4_french.svg par les auteurs Just plain Bill et Romainbehar.

Ici l'emprunt en *Sol Majeur* donne une fonction de tonique à l'accord du premier temps de la mesure 7, ce qui donne une sensation de déplacement de centre tonal. On appelle ce bref emprunt à une autre tonalité, une *tonicisation*. Ce terme vient de l'allemand *Tonikalisierung* [119]. Plus particulièrement, on emploie le terme *dominante secondaire* pour caractériser un accord de dominante de la tonalité empruntée.

Ainsi, à chaque moment de la partition, on pourra associer deux tonalités, celle de la modulation et celle de la tonicisation⁵. Les tonalités locales, modulations et tonicisations se trouvent à des moments précis de la partition : il est ainsi possible de les modéliser, les annoter dans des corpus (chapitre 4), et proposer des méthodes informatiques d'annotation (semi-)automatique (chapitre 5).

La tonalité constitue ainsi une sorte de "grammaire" musicale grâce à laquelle le compositeur forme des phrases. Ces phrases sont conclues par des cadences qui sont présentées à la section suivante.

2.3 Un processus de terminaison des phrases musicales : la cadence

2.3.1 Principes fondamentaux

La musique est organisée en unités structurales fonctionnelles [17]. Dans la musique tonale occidentale, ces unités se terminent généralement par des formules harmoniques à fonc-

5. M. Rigaudière parle de modulation forte et de modulation passagère [110, p. 161]

tion stabilisatrice appelées *cadences*. Le mot cadence vient lui même du latin *cadere* qui signifie "tomber". Par analogie avec la littérature, une cadence peut être considérée comme une ponctuation musicale. On rencontrera ainsi des formules harmoniques suspensives, telle les *virgules* des phrases, appelées *demi-cadences* mais aussi des formules harmoniques très conclusives appelées *cadences parfaites* et apparentées à des *points*. Une terminologie des différents types de cadences est présentée à la section 2.3.2.

Malgré leur fonction structurelle bien définie, la définition de ce qu'est une cadence reste ambiguë. A. Blombach s'appuie sur une dizaine d'ouvrages de théorie musicale pour proposer une définition de la cadence comme étant *tout élément musical ou combinaison d'éléments musicaux, y compris les silences, qui indiquent une détente ou une terminaison relative de la musique* [16].

Selon la définition de W. E. Caplin qui est celle que nous adoptons pour ce manuscrit, la cadence agit comme un processus de terminaison formelle d'une progression harmonique et d'une progression mélodique dont le but est le *point d'arrivée cadentiel*. Ce qui suit la cadence sera considéré comme *postcadentiel* si on n'ouvre pas un nouveau processus formel [19].

L'étude du phénomène cadentiel est encore très vivace, en témoigne la parution en 2015 d'un ouvrage collectif qui lui est dédié [96].

2.3.2 Types de cadences

On distingue deux types de cadences, les cadences *conclusives* et les cadences *suspensives*.

Les cadences conclusives sont caractérisées par un point d'arrivée cadentiel sur une harmonie de tonique (I) ou au moins une volonté d'arrivée sur cette harmonie.

La cadence la plus stable et la plus conclusive est la cadence *parfaite*. Elle est caractérisée par une progression harmonique V vers I, les deux accords étant en position fondamentale et la voix portant la mélodie arrivant sur la tonique. On l'appelle *perfect authentic cadence* en anglais et on la note PAC. Un exemple de PAC est présenté à la figure 2.4.

The image shows a musical score for the 18th Sonata for Piano, Opus 31 No. 3 by Ludwig van Beethoven, measures 10 to 17. The score is in 3/4 time, key of B-flat major (Mi bémol Majeur), with a tempo of 118. The music shows a progression from a II chord to a V/V chord, then to a V chord, and finally to an I chord. The final I chord is highlighted with a red vertical bar. Performance markings include 'p' (piano), 'Ritard.' (ritardando), 'cresc.' (crescendo), and 'sf' (sforzando). The tempo changes to 'a Tempo.' after measure 15.

FIGURE 2.4 – 18^e Sonate pour piano opus 31 no.3 de L. V. Beethoven mesures 10 à 17. Exemple d'une cadence parfaite dont le point d'arrivée est encadré en rouge au premier temps de la mesure 17. La progression harmonique est typique de la cadence parfaite II-V-I. L'ajout du V/V apporte un peu plus de tension à cette préparation. La résolution de toutes les tensions sur la tonique est ainsi particulièrement appréciable. Notons également le changement de texture tout de suite après la cadence parfaite, indiquant qu'on entre dans une nouvelle unité structurelle.

La cadence *imparfaite* est un peu moins stable. Elle observe la même progression harmonique que la cadence parfaite mais les accords ne sont pas forcément en position fondamentale

et la mélodie ne s'arrête pas forcément sur la tonique. Comme elle est moins conclusive, elle peut-être considérée comme une conclusion partielle et être suivie par une cadence parfaite (elle l'est souvent d'ailleurs dans les quatuors de W. A. Mozart). On l'appelle *imperfect authentic cadence* en anglais et on la note IAC.

Pour certaines cadences imparfaites, les deux harmonies finales sont en position fondamentale (*root position*, comme pour une cadence parfaite) et la mélodie aboutit plutôt sur la tierce (et extrêmement rarement la quinte) de l'accord de tonique (ce qui en fait une cadence imparfaite). Ces cadences sont presque aussi conclusives que des PACs et se rencontrent fréquemment dans la littérature Mozartienne. Elles sont notées par l'équipe Algomus *rIAC* (pour *root position Imperfect Authentic Cadence*) dans l'article [55], Notons cependant que certains théoriciens français classent ce type de cadence comme une cadence parfaite (voir par exemple [2]).

Amener une cadence est un processus qui comprend une préparation et une résolution [19]. Il arrive parfois que le compositeur trompe l'auditeur en faisant entendre une autre harmonie finale que celle de tonique, bien souvent le VI^e degré. Cette cadence est appelée *cadence rompue* ou *deceptive cadence* en anglais. On la note DC. Un exemple de ce type de cadence est présenté à la figure 2.5.

FIGURE 2.5 – 3^e mouvement issu de la 4^e Sonate pour Piano K. 282 (159g) de W. A. Mozart, mesures 28 à 35. Exemple d'une cadence rompue dont le point d'arrivée cadenciel est sur le premier temps de la mesure 31. On entend la progression harmonique de la cadence parfaite (I-ii-V) de la mesure 28 à la mesure 30 mais au lieu d'aller sur la tonique à la mesure 31 et fermer la phrase, Mozart choisit de faire entendre le VI^e degré ce qui crée la cadence rompue. Ce n'est que partie remise car ces mesures sont répétées aux mesures 32 à 35 et cette fois-ci on a bien le I^e degré attendu.

La *cadence évitée* ou *cadence élidée* est une cadence rompue particulière qui n'est pas théorisée par tout le monde [121]. La cadence évitée est une cadence parfaite dont la résolution est remplacée par une répétition de la préparation de la cadence. On la rencontre assez souvent chez Mozart, particulièrement dans les secondes zones thématiques des œuvres en forme sonate. On la note EV.

Il n’y a qu’une cadence suspensive, la *demi-cadence*. Elle s’achève sur une harmonie de dominante, en général en position fondamentale. On l’appelle *half cadence* en anglais et on la note HC. Un exemple de ce type de cadence est présenté à la figure 2.6.

The image shows a musical score for Schubert's Impromptu op. 142, measures 1 to 4. The score is in 3/4 time, Si bémol Majeur, Andante, and piano (p). The harmonic progression is I - ii - V/V - V. A red vertical bar highlights the final chord (V) in measure 4, labeled as V and HC (Half Cadence).

FIGURE 2.6 – 3^e *impromptu* issu des *Impromptus* op. 142 de F. Schubert, mesures 1 à 4. Exemple d’une demi-cadence dont le point d’arrivée cadentiel est sur le deuxième temps de la mesure 4. La progression harmonique I-ii-V/V-V est typique de la demi-cadence mais n’en est pas caractéristique, l’harmonie de dominante de dominante est loin d’être une obligation pour la demi-cadence. Cette petite phrase est reprise aux mesures 5 à 8 (non affichées) et se conclue cette fois par une cadence parfaite. Dans ce cas, on parle d’antécédent pour la phrase des mesures 1 à 4 et de conséquent pour la phrase des mesures 5 à 8.

Il existe enfin une dernière cadence, la *cadence plagale* dont le statut de cadence est remis en cause par certains théoriciens. Il s’agit d’une progression IV vers I qui doit plus être vue comme une progression post-cadentielle qu’une cadence en tant que telle [19].

Chaque type de cadence apporte une sensation de conclusion plus ou moins prononcée. La cadence la plus forte est la PAC suivie par la rIAC, l’IAC, la HC, la DC et enfin l’EV [122].

Avant un accord de dominante, on entend souvent un ou plusieurs accords de préparation. La cadence parfaite type, telle que la décrit W. E. Caplin, suit la progression harmonique I₆, ii₅⁶ (ou tout autre accord ayant une fonction de *sous-dominante*), V, I. La dominante peut être séparée en deux harmonies, la sixte et quarte de I (aussi appelée sixte et quarte de cadence) notée V₄⁶ suivie de l’harmonie de V à proprement parler. Ce genre de préparation très codifiée et convenue donne la plus forte sensation de terminaison. C’est pourquoi une cadence rompue ou évitée crée de la tension en permettant une extension de la phrase musicale, retardant la résolution attendue de quelques mesures.

Toutes les cadences marquent la fin des phrases musicales. Elles sont donc des événements privilégiés à observer car elle agissent comme frontière entre différentes sections de la forme sonate. On présente la forme sonate dans la section suivante.

2.4 Structurer la musique de façon dialectique : la forme sonate

Nous présentons maintenant la forme sonate. Cette thèse s’inscrit dans le défi que relève l’équipe Algomus de parvenir à l’analyse automatisée de cette forme. Cette présentation s’inspire en partie de l’article que nous avons publié sur la détection des formes sonates [7].

2.4.1 La forme sonate, une organisation tonale et thématique

La *forme sonate* est une formalisation à posteriori de pratiques largement utilisées par les compositeurs depuis le milieu du XVIII^e siècle. Elle peut être vue comme une évolution de l'aria et de la forme du concerto baroque [115, 65].

La forme sonate organise la partition en associant étroitement relations de tonalités et relations thématiques. Elle est construite sur un contraste entre deux zones tonales thématiques, notées *P* et *S*, présentées dans une première grande section appelée *Exposition*. *P* est dans la tonalité principale, *S* est dans une autre tonalité, la tonalité secondaire. Les thèmes de *P* et de *S* peuvent être soit identiques ou très proches (sonate monothématique, surtout chez J. Haydn), soit différents (sonate bithématique). Outre les différences harmoniques et (pour les sonates bithématiques) mélodiques, *P* et *S* contrastent fréquemment par leur caractère.

La différence de tonalité crée une dissonance à grande échelle (d'après les termes de C. Rosen [115]) qui est amplifiée dans une seconde grande section, le *Développement*. Cette dissonance est résolue dans une troisième grande section, la *Réexposition*, dans laquelle les deux zones tonales sont revisitées dans la tonalité principale. On les notera alors *P'* et *S'*. Le découpage de la forme sonate du 2^e mouvement du quatuor K428 de W. A. Mozart est présenté à la figure 2.7.

Quelques repères historiques. Bien qu'on en trouve des embryons chez C. P. E. Bach ou Scarlatti, la forme sonate fut élaborée essentiellement par J. Haydn, W. A. Mozart et L. V. Beethoven. On la trouve généralement dans les premiers mouvements des œuvres de genre sonate, symphonie, concerto ou musique de chambre, mais on peut aussi la rencontrer dans les autres mouvements, un peu modifiée (sonate sans développement, rondo-sonate ou même esprit de la sonate dans des formes menuet).

Les théories à propos de la forme sonate classique émergent près de cinquante ans plus tard, au cours des années 1820 [108, 83, 36]. Une des premières formalisations de cette forme est la *grande coupe binaire* proposée par A. Reicha, 30 ans après la mort de Mozart [108]. Tout comme la tonalité, la forme sonate perdurera durant la période romantique et sera infléchi par les compositeurs qui la transformeront de l'intérieur. On peut notamment citer la *Phantasie*, op. 17 de R. Schumann, qui croise la forme sonate avec la fantaisie ou encore la *Sonate pour piano en Si mineur* de F. Liszt, qui peut être analysée comme une grande forme sonate bithématique ou comme les quatre mouvements d'une œuvre du genre sonate [127].

2.4.2 Théories récentes sur la forme sonate

De nos jours, la forme sonate est toujours enseignée dans le cadre des classes d'analyse musicale, d'histoire de la musique ou de composition. Nous avons vu ces dernières décennies un regain d'intérêt des théoriciens de la musique pour l'analyse de la musique de la période classique [106, 115, 64, 17, 18, 63, 77, 88, 65, 57, 58]. De nouvelles théories plus exhaustives émergent, en particulier pour l'analyse de la forme. Nous présentons maintenant deux de ces théories.

W. E. Caplin, dans l'ouvrage *Classical Form* [17], revisite l'analyse schoenbergienne de la forme en proposant une théorie des fonctions formelles dans laquelle il combine segmentation de la partition et analyse formelle. La partition est segmentée en unités élémentaires appelées "idées de base" ou "idées contrastantes". Celles-ci sont combinées en unités formelles au ni-

The figure displays a musical score for the 2nd movement of Mozart's K. 428, annotated with tonal and formal labels. The score is divided into sections: Exposition (I:HC MC), Development (Dev), and Recapitulation (I:HC MC', I:PAC ESC). Labels include P, TR, S, C, S', C', Dev, and RT, indicating thematic zones and transitions.

FIGURE 2.7 – Andante con moto du 2^e mouvement du 16^e quatuor à cordes K. 428 composé par W. A. Mozart. Figure extraite de notre article [7], mouvement encodé en Lilypond par M. Tomasi pour le projet Mutopia. Ce mouvement lent est en forme sonate. Selon les notations de J. Hepokoski [65], les premières zones tonales thématiques (P/P') sont suivies par les transitions (TR/TR'), qui sont conclues par les césures médianes (MC/MC') – des demi-cadences (HC) dans la tonalité principale Mi \flat Majeur (I). La zone tonale thématique secondaire (S) et la zone conclusive (C) sont dans la tonalité secondaire Si \flat Majeur (V). S' et C', les versions de S et C dans la réexposition sont dans la tonalité principale. La zone thématique S est conclue par une PAC dans la tonalité secondaire qui est aussi l'EEC. Sa réexposition S' s'achève par une PAC dans la tonalité principale qui est aussi l'ESC. Entre l'Exposition et la Réexposition, on trouve le Développement (Dev) qui est le lieu de plusieurs modulations dans d'autres tonalités. Il est conclu par une retransition (RT) mettant en valeur la dominante de la tonalité principale afin de préparer la Réexposition.

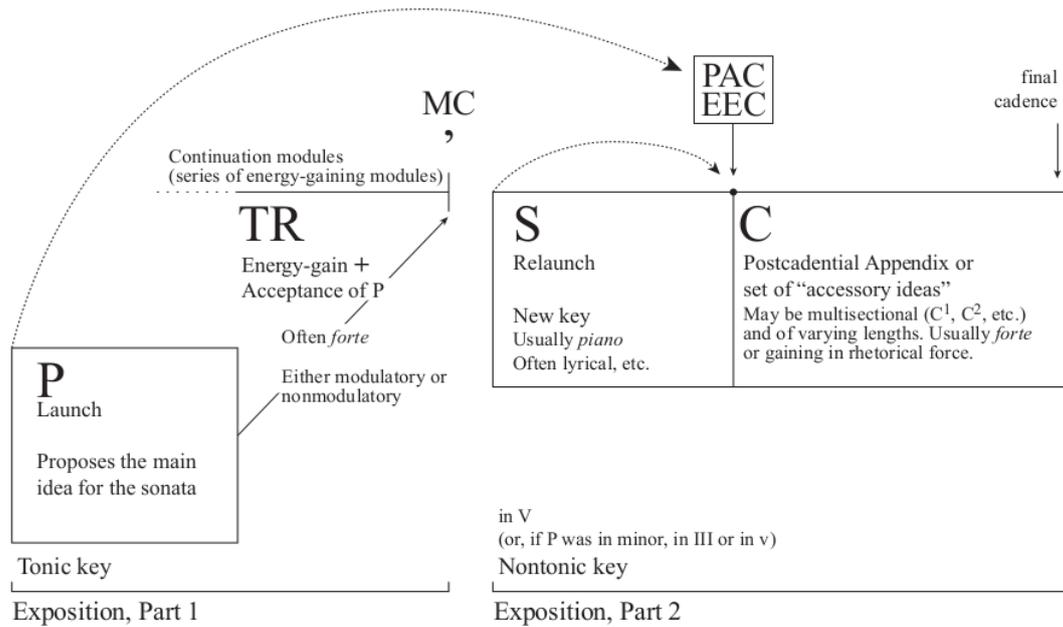


FIGURE 2.8 – Structure de l'exposition à deux parties dans les formes sonates. Figure extraite de l'ouvrage de J. Hepokoski et W. Darcy [65].

veau de la phrase⁶ (comme l'antécédent, le conséquent, la présentation) qui sont elles-même combinées en unités formelles au niveau thématique. Ces unités thématiques peuvent être des "sentences", des "periods" ou des "small ternary". Cette approche qui part de la base permet de construire la forme sonate dans son intégralité jusqu'à arriver aux trois grandes sections *Exposition, Développement et Réexposition*, avec inclusion potentielle de deux autres fonctions, *Introduction* et *Coda*.

J. Hepokoski et W. Darcy introduisent quant à eux la *Sonata Theory* [65]. La forme sonate y est considérée comme "un système ordonné d'options génériques disponibles" permettant l'extension temporelle de la partition (*ibid.*, p. 15). La figure 2.8 présente un schéma de l'organisation de l'exposition de la forme sonate telle qu'elle est définie par J. Hepokoski et W. Darcy. On remarque que l'exposition est divisée en deux parties. La première contient la première zone thématique *P* ainsi qu'une transition *TR* dont le but est de gagner de l'énergie pour permettre la modulation dans la tonalité secondaire et préparer un espace musical permettant la présentation du second thème. Tout le processus mené dans la première partie a un but cadentiel appelée la césure médiane *MC*. Celle-ci est un moment de rupture dans la musique marquant la fin de la première partie de l'exposition et le début de la seconde.

La deuxième partie de l'exposition contient la seconde zone thématique *S* dans la tonalité secondaire. Celle-ci s'achève par la "terminaison expositionnelle essentielle" *EEC* qui est une cadence parfaite dans la tonalité secondaire. L'*EEC* est le but cadentiel des deux parties de l'exposition. Elle est suivie par une zone conclusive *C* qui est en général une succession de formules cadentielles et qui peut éventuellement préparer le développement et la reprise de l'exposition. L'*EEC* crée une dissonance qui est résolue dans la réexposition par la terminaison

6. La phrase ici constitue une unité temporelle de 4 mesures.

structurelle essentielle *ESC*, une cadence parfaite dans la tonalité principale qui clôt *S'*. L'*ESC* est le troisième et dernier but cadentiel de la forme sonate.

Les deux théories de W. E. Caplin et de J. Hepokoski / W. Darcy sont confrontées dans l'ouvrage *Musical Form, Forms and Formenlehre* [20] ainsi que celle de J. Webster, par un jeu de présentations et de commentaires. Nous avons décidé, pour cette thèse, de nous appuyer sur la théorie de J. Hepokoski et de W. Darcy pour modéliser informatiquement la forme sonate. En effet, cette théorie caractérise de façon exhaustive chacune des sections de la forme sonate, ce qui permet de lever beaucoup d'imprécisions sur la *position des frontières* dans la partition musicale. Ce choix se retrouve dans les annotations de corpus de forme sonate (section 4.3), mais surtout dans la manière de considérer certains éléments. Nous nous intéressons particulièrement à la *césure médiane* de par son rôle central dans la définition de la forme sonate.

2.4.3 Préparation et articulation de la césure médiane

La MC conclut un processus de gain d'énergie initié durant TR. Elle ouvre ainsi un espace pour S grâce à cette énergie gagnée. Pour reprendre l'analogie présentée par J. Hepokoski et W. Darcy, la césure médiane est le moment de l'ouverture des portes d'un ascenseur après avoir atteint un étage supérieur [65].

Nous présentons deux MC aux figures 2.9 et 2.10. Le processus amenant à la MC est présenté dans l'article [64]. Il commence lorsque la musique arrive sur une *dominante structurelle*. Cette dominante peut être *entretenu* par des broderies autour des notes de l'accord de dominante, la répétition et la tenue de la note pédale de dominante. La nuance y sera en général *forte* afin de conserver l'énergie. À la fin de ce processus, il peut y avoir la présence de plusieurs *coups de marteau* (en général deux ou trois) qui consistent à marteler cet accord de dominante sous forme fondamentale.

Vient alors l'articulation de la césure à proprement parler, qui peut être soit une pause à toutes les voix (appelée *general pause*) ou une transition vers S sous forme de remplissage (*caesura-fill*) afin de faire le pont entre les deux zones.

Bien que sur le papier, ces césures médianes semblent bien définies, dans la réalité, il est difficile de trouver 2 MC qui se ressemblent, à commencer par celles données en exemple.

J. Hepokoski et W. Darcy classent les MC en 4 catégories suivant la cadence qui les caractérise et leur position dans l'exposition :

- les MC du premier type sont associées à des demi-cadences dans la tonalité secondaire⁷. Ces MC se retrouvent en général entre 25% et 50% voire 60% de l'exposition.
- les MC du deuxième type sont associées à des demi-cadences dans la tonalité principale. Elles se retrouvent entre 15% et 45% de l'exposition.
- les MC du troisième type sont associées à des cadences parfaites dans la tonalité secondaire. Elles se retrouvent entre 50% et 70%, voire 75% de l'exposition.
- les MC du quatrième type sont associées à des PAC ou IAC dans la tonalité principale. Il s'agit très souvent d'une MC qui est aussi la cadence de la fin de P, dans ce cas S suit directement P.

On peut aussi caractériser la césure médiane comme étant ce qui précède directement S. Cette zone est thématique et on peut légitimement penser que c'est plus facile à trouver sur la

7. La tonalité secondaire est en général le ton majeur de la dominante pour une tonalité principale majeure ou le ton mineur de la dominante ou le relatif majeur pour une tonalité principale mineure.

FIGURE 2.9 – 1^{er} mouvement de la Sonate pour piano en La mineur, K. 310 de W. A. Mozart, mesures 15–24. Le point d’arrivée de la demi-cadence est sur le premier temps de la mesure 16 (V arrivée entouré en orange). Les mesures 16 à 21 ont une fonction prolongationnelle. Elles sont contruites sur une pédale de dominante (Pedal V en bleu) et une oscillation harmonique de plus en plus rapide entre la dominante et la tonique de Do mineur. La tension de dominante est renforcée par le contraste forte/piano/forte aux mesures 16, 18 et 20. Cela mène au triple coup de marteau à la mesure 22 (THB en vert) et à la véritable articulation de la MC sur le 4^{ème} temps de la mesure 22 (MC en rouge). La caesura-fill juste après (en bleu plus clair) amène à S et à sa nouvelle unité thématique.

FIGURE 2.10 – 1^{er} mouvement du Quatuor à cordes en Ré mineur, K. 421 de W. A. Mozart, mesures 22–24. Cette MC est faible : pas d’arrivée de dominante, pas d’arrêt sur le degré V, pas de coups de marteau, et plutôt une déperdition d’énergie qu’un gain. Cette faiblesse de la MC peut être due à un refus de césure médiane de premier type sur le 3^{ème} temps de la mesure 14, non montrée sur cette figure (la dominante arrive mesure 12). Le compositeur retarde ainsi l’arrivée de S et continue sur TR, peut-être pour créer une surprise.

partition. M. Richards propose sept signaux dont la présence peut aider à trouver le début de S [109]. Ceux-ci sont :

- la présence d'une harmonie de tonique de la nouvelle tonalité,
- une phrase ayant une fonction de commencement,
- juste avant le début de S, la présence d'un accord de fin de phrase,
- une interruption textuelle (c'est-à-dire la césure médiane),
- des changements dans la texture,
- des changements de nuances,
- un matériau mélodique caractéristique.

Si tous ces signaux sont plutôt complexes à formaliser informatiquement, nous en retenons certains, comme les changements de texture ou la présence d'accords particuliers, comme outils potentiels pour modéliser la césure médiane.

2.5 Bilan

Tonalité, cadences et forme sonate sont des notions qui dépassent le simple cadre local ou harmonique. Elles contribuent à l'élaboration d'un système complet et complexe, tel un langage, ainsi qu'à l'organisation du discours musical. La suite de ce manuscrit a pour but de proposer des modélisations de ces concepts haut-niveau. Le prochain chapitre présente quelques travaux en informatique musicale à ce propos.

Chapitre 3

État de l'art en musicologie numérique sur l'estimation de la tonalité et sur la détection des cadences

Contenu

3.1 Estimation automatisée de la tonalité	28
3.1.1 L'algorithme de Longuet-Higgins et Steedman	30
3.1.2 Carol L. Krumhansl et les <i>profils tonaux (pitch profiles)</i>	30
3.1.3 Monopole des profils tonaux pour l'estimation de tonalité	32
3.1.4 Autres méthodes pour l'estimation de tonalités	35
3.2 Perception, modélisation et détection automatisée des cadences	37
3.2.1 Cadences et perception de fermeture	38
3.2.2 Modélisation des cadences pour la prédiction automatisée	39
3.3 Bilan	40

Dans ce chapitre, nous présentons les travaux ayant cherché à estimer les tonalités (Section 3.1) et détecter les points d'arrivée cadentiels (Section 3.2). Ces travaux en informatique musicale appréhendent et modélisent des notions haut-niveau, la plupart étant, comme l'ensemble de cette étude, sur des données symboliques (partitions). Ils peuvent s'appuyer sur des travaux de perception et de musicologie.

3.1 Estimation automatisée de la tonalité

Ressentir la tonalité d'une œuvre musicale – au sens de tonalité locale/globale comme évoqué en section 2.2 – est l'une des premières actions de l'analyste, mais aussi de l'interprète, lorsqu'il découvre une partition. Très tôt dans l'histoire de l'informatique musicale, dès les années 70, des auteurs ont proposé des algorithmes visant à déterminer la tonalité d'une œuvre. Au moins une vingtaine d'études s'attaquent à ces questions (Table 3.1).

Les premiers algorithmes d'estimation des tonalités se sont essentiellement basés sur les hauteurs de notes. En effet, on associe à chaque tonalité, une gamme pour les tonalités majeures et trois gammes pour les tonalités mineures. On peut donc estimer la tonalité globale d'une pièce en appliquant la *règle* suivante : la tonalité globale d'une pièce est celle dont les notes de la gamme sont présentes au début du morceau. On peut y adjoindre d'autres règles. Cette approche est étudiée dans [80] et [135] (Section 3.1.1).

Les approches postérieures utilisent les *profils tonaux*. Ceux-ci *pondèrent les notes* en fonction de leur appartenance à une tonalité donnée, en particulier la tonique (I, le degré sensé commencer et finir une partition [105]) et la dominante (V, celui contenant la sensible (dissonance majeure) et précédant directement I dans les cadences parfaites [105]). Les profils tonaux quantifient la hiérarchisation entre les hauteurs de notes. Ils pourront être constitués aussi bien d'après le résultat d'expériences cognitives [75], définis arbitrairement par des musicologues [129] ou calculés d'après la distribution des notes sur des corpus d'étude [1] (Sections 3.1.2 et 3.1.3).

Les algorithmes utilisant les profils tonaux calculent une *corrélation* entre un profil tonal d'une tonalité donnée et les notes sur la partition. La corrélation n'est pas l'exclusivité des profils tonaux et peut aussi être calculée pour d'autres types de modèles comme les modèles par *apprentissage* de type réseau de neurones, par exemple.

Les approches contemporaines de cette thèse, en particulier celles visant à estimer la tonalité locale, s'intéressent désormais au *contexte temporel* des progressions harmoniques. En effet, le système tonal est harmoniquement dirigé [90] et il faut prendre en compte un accord et son passé proche pour pouvoir prendre une décision quant à la tonalité d'un extrait musical. Par exemple, l'enchaînement d'un accord de *Fa Majeur* vers un accord de *Sib Majeur* est plus significatif de la tonalité de *Sib Majeur*¹ que de celle de *Fa Majeur* alors que pourtant ces accords et donc ces notes appartiennent aux deux tonalités. Cette approche sera aussi développée par la suite dans le chapitre 5.

Nous présentons également brièvement un panel de travaux sur l'estimation de tonalité locale à la section (Section 3.1.4). Ceux-ci utilisent les méthodes présentées ci-dessus, des méthodes complémentaires ou une combinaison de ces méthodes pour l'estimation de tonalité.

1. Notamment grâce à la résolution de la tension créée par la sensible sur la tonique.

	Symbolique	Audio	Méthode	Règles	Corrélation	Pondération	Contexte temporel	Apprentissage	Corpus de test	Global	Local	Résultats
Longuet-Higgins 1971	X		Elimination	X					48 fugues (Bach, CBT)	X		100,00 % (a)
Krumhansl 1990	X		Pitch profiles		X	X			48 préludes (Bach, CBT)	X		91,70 % (b)
									24 préludes de Chostakovitch	X		70,80 %
									24 préludes de Chopin	X		45,80 %
									48 fugues du CBT	X		91,70 %
									24 fugues de Chostakovitch	X		91,70 %
									490 œuvres de la BD Humdrum	X		74,20 %
Vos 1996	X		Elimination	X					48 fugues du CBT	X		81,30 %
Temperley 1999	X		Pitch profiles		X	X			48 fugues du CBT	X		87,50 %
									Kostka Payne (pas d'inventaire)		X	83,80 %
									490 œuvres de la BD Humdrum	X		88,60 %
Purwins 2000	X		Pitch profiles		X	X			(1 prélude de Chopin)		X	
Chew 2002	X		Spiral array		X	X			(2 œuvres de Bach)		X	
Aarden 2003	X		Pitch profiles		X	X		X	48 fugues du CBT	X		95,80 %
									48 préludes du CBT	X		95,80 %
									24 préludes de Chopin	X		91,70 %
									1000 Essen folksongs	X		93,50 % (c)
									250 exemples de MuseData	X		93,40 % (c)
									490 œuvres de la BD Humdrum	X		91,70 %
Chai 2005	X		HMM		X	X			10 pièces pour piano		X	~ 85 % (d)
Bellmann 2006	X		Pitch profiles		X	X		X	48 fugues du CBT	X		93,80 %
									48 préludes du CBT	X		95,80 %
									24 préludes de Chopin	X		87,50 %
									490 œuvres BD Humdrum	X		91,10 %
Catteau 2006	X		Pitch profiles		X	X			10 séquences d'accords		X	86,70 %
Izmirlı 2007	X		Non negative matrix factorization, corrélation Pitch profiles		X	X			17 morceaux pop	X		64,70 %
									17 morceaux pop		X	71,80 %
									152 pièces du Naxos Set	X		80,90 %
									152 pièces du Naxos Set		X	77,10 %
									17 extraits de Kostka-Payne	X		94,10 %
									17 extraits de Kostka-Payne		X	67,60 %
Papadopoulos 2009	X		Pitch profiles / HMM		X	X	X		5 mouvements des sonates pour piano de Mozart		X	80,20 %
Rocher 2010	X		Dynamic programming key profile		X	X	X		174 chansons des Beatles		X	62,40 %
Mearns 2011	X		HMM sur les accords Chord profiles		X	X	X		12 Chorals de Bach		X	79,00 %
Sapp 2011	X		Pitch profiles		X	X			segments de la BD MuseData	X	X	97,90 %
									48 préludes du CBT	X		97,90 %
									24 préludes de Chopin	X		91,70 %
									490 œuvres de la BD Humdrum	X		90,40 %
Albrecht 2013	X		Pitch profiles		X	X		X	490 œuvres de la BD Humdrum	X		90,00 %
Pauwels 2014	X		Modèle probabiliste acoustique infos musicologiques		X				142 extraits du SEMA set		X	63,60 %
									210 extraits de MIREX2009		X	74,90 %
Weiss 2014	X		Pitch profiles		X	X			Exemples variés			
Chen 2018	X		Réseaux de neurones		X			X	6 mouvements de sonates pour piano de Beethoven		X	66,70 %
2019	X		Réseaux de neurones		X			X	32 mouvements de sonates pour piano de Beethoven		X	78,40 %
Napoles Lopez 2019	X		Hmm + pitch profiles		X	X	X		(1 prélude de Chopin)	X	X	
Micchi 2020	X		Réseaux de neurones		X			X	10 thèmes et variations Mozart		X	82,90 %
									17 themes et variations Beethoven			
									70 mouvements de quatuors à cordes Beethoven (ABC)			
									32 mouvements de sonates pour piano Beethoven (BPS-FH)			
									24 Préludes de Bach			
									48 chansons romantiques			

TABLE 3.1 – Méthodes pour l'estimation de la tonalité, discutées ou citées dans ce chapitre. Pour chaque méthode est indiquée le type de fichier en entrée (audio, symbolique), le type de méthodes utilisées (basée sur des règles, basée sur une corrélation calculée entre les notes des gammes des tonalités et la partition, une corrélation avec pondération des notes des gammes, une prise en compte du contexte temporel autour du temps évalué ou basée sur un apprentissage statistique ou autre). Les résultats obtenus, issus des articles correspondants, sont uniquement indicatifs. Les méthodes d'évaluation sont trop différentes pour que les modèles soient vraiment comparables. L'évaluation des algorithmes d'estimation de tonalité est une des questions étudiées par la thèse à la section 5.4.

(a) Règles très (trop) complaisantes qui expliquent le résultat.

(b) Évaluation sur les 4 premières notes de l'œuvre uniquement.

(c) Ensemble du corpus en tonalité majeure non-modulante.

(d) Résultats détaillés non précisés dans l'article.

3.1.1 L'algorithme de Longuet-Higgins et Steedman

L'un des premiers algorithmes est proposé par H. C. Longuet-Higgins et M. J. Steedman en 1971 dans l'article *On interpreting Bach* [80]. Dans cet article, les auteurs conçoivent deux programmes, à base de règles, l'un pour déterminer la métrique et l'autre pour trouver les relations harmoniques entre les notes des sujets de fugues de Bach. Une fugue commence en effet par l'exposition du *sujet* à une seule voix, et ce sujet, à lui tout seul, permet d'établir la tonalité.

Ce second programme se déroule en plusieurs étapes dont la première est la détermination de la tonalité des différents sujets. Partant du constat qu'un auditeur est de plus en plus familier avec un sujet de fugue au fur et à mesure que celui-ci est joué, les auteurs proposent un programme qui aura une meilleure conscience de la tonalité et de la métrique lors de son parcours du sujet. Ainsi au début du sujet, on n'a aucune idée de quelle pourra être la tonalité (toutes sont considérées) et l'estimation sera de plus en plus précise.

Concrètement, on procède par élimination. Chaque nouvelle note dans le sujet est insérée dans la liste des notes déjà entendues, puis on élimine toutes les tonalités qui ne contiennent pas les notes de cette liste. Lorsqu'il ne reste qu'une tonalité, c'est que c'est celle du sujet. S'il en reste plusieurs, on applique la règle de "préférence tonique-dominante". On prend la tonalité dont la première note du sujet est la tonique si elle fait partie des tonalités candidates, sinon celle dont la première note du sujet est la dominante.

L'algorithme trouve correctement la tonalité de tous les sujets de fugue mais il convient de remarquer que la règle de "préférence tonique-dominante" est beaucoup trop forte (elle force la décision pour 17 des 48 sujets) et fonctionne uniquement car les auteurs travaillent sur des sujets de fugues.

Malgré ses résultats, cet algorithme ne fut que très peu utilisé à notre connaissance. Seul un article de P. G. Vos et de E. W. Van Geenen [135] en 1996 présente une approche similaire bien que plus complexe et complète puisque le modèle proposé incorpore aussi des informations sur les accords et leur fonction. Si cet algorithme n'a pas suscité plus d'engouement, c'est parce qu'à la fin de la décennie 1970-1979, une nouvelle façon de penser la gamme fait son apparition, le *profil tonal*.

3.1.2 Carol L. Krumhansl et les *profils tonaux (pitch profiles)*

Carol L. Krumhansl, psychologue et cognitiviste de la musique, a marqué notre domaine par ses études sur la tonalité. En 1979, elle conduit avec Roger N. Shepard deux premières expériences en perception musicale, décrites dans l'article *Quantification of the Hierarchy of Tonal Functions within a Diatonic Context* [76].

Afin d'estimer à quel point chaque note est reliée à une gamme, ils conduisent une première expérience sur 24 personnes, classées en 4 groupes suivant leur connaissance de la théorie musicale occidentale (aucune, niveau moyen (en moyenne 5.5 années d'études), très bon niveau (7.4 années d'études) et niveau exceptionnel (oreille absolue, 17 années d'études)). On joue la gamme de Do Majeur ascendante ou descendante (sans la répétition de la tonique) puis une note au hasard. Le participant doit estimer à quel point la note complète la gamme, sur une échelle de 1 (pas du tout) à 7 (complètement).

La deuxième expérience ne concerne que les personnes des groupes de très bon niveau et de niveau exceptionnel. Des notes altérées par des quarts de tons sont ajoutées au stimulus.

Les figures correspondant aux résultats de ces expériences prennent la forme de courbes qui présentent la hauteur de note en abscisse et la note moyenne des sujets en ordonnée (Figure 3.1). Le *profil tonal* pour la gamme de Do Majeur est né. Le profil tonal obtenu dans la deuxième expérience est sensiblement le même que pour le groupe de très bon niveau de la première expérience.

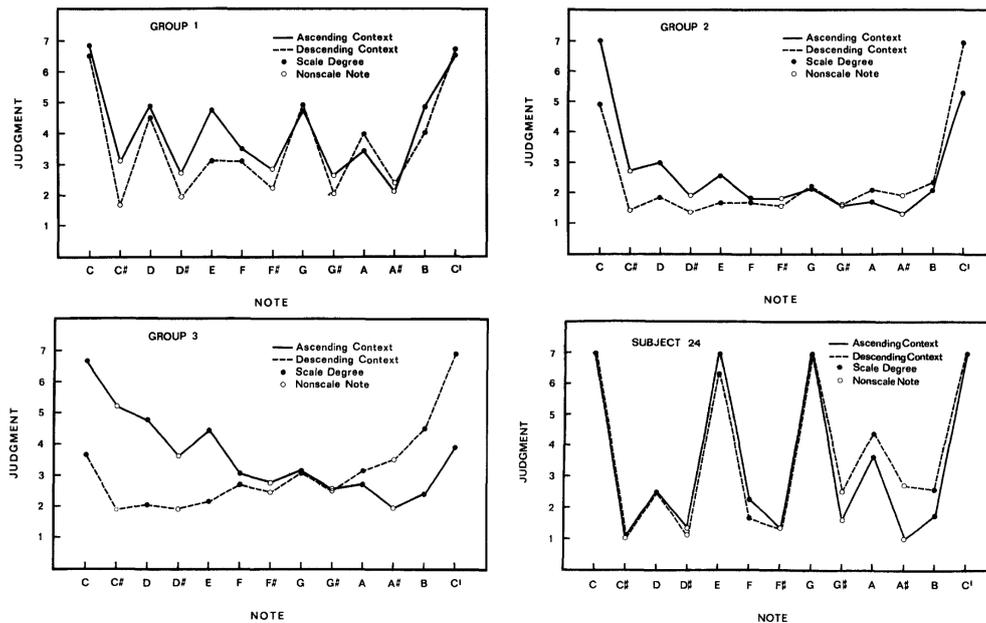


FIGURE 3.1 – Collage de figure issues de Quantification of the Hierarchy of Tonal Functions Within a Diatonic Context [76]. Chacun des graphiques correspond au jugement de l'appartenance de la note à une gamme pour un des groupes avec en abscisse la note considérée et en ordonnée la moyenne de l'évaluation proposée par les participants. Le groupe 1 correspond au groupe de très bon niveau de théorie musicale occidentale. Le groupe 2 correspond au niveau moyen. Le groupe 3 au groupe peu familier et le groupe 4 (ou sujet 24) correspond à la personne ayant un niveau exceptionnel ainsi que l'oreille absolue.

Les résultats de l'expérience montrent que plus la personne a une éducation musicale conséquente, plus le profil tonal met en évidence les notes importantes de la gamme, créant ainsi une hiérarchie entre les différents degrés d'une tonalité. Les notes les plus importantes pour le groupe de très bon niveau sont la tonique *Do* ainsi que les notes complétant l'accord de tonique (*Mi* et *Sol*). Puis viennent la sensible *Si* et les notes complétant la gamme de *Do* Majeur (*Ré*, *Fa* et *La*). Le profil tonal pondère donc l'importance des notes d'une gamme pour une tonalité donnée.

Pour le groupe ayant la culture musicale occidentale la moins développée, seul le *Do* semble important pour conclure la gamme ascendante ou descendante (le reste du profil est plat), alors que pour la personne ayant un niveau exceptionnel, les trois notes de l'accord parfait majeur ont une valeur maximale. Ce résultat semble mettre en évidence que le sens de l'harmonie est plus développé chez les personnes pratiquant la musique tonale occidentale, puisqu'ils ne s'attendent pas seulement à une note, mais à une harmonie complète.

En 1982, C. L. Krumhansl continue ses travaux avec l'aide de E. J. Kessler et ensemble, ils publient l'article *Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys* [75]. Ils reproduisent la première expérience de 1979 en ne se limitant pas à Do Majeur mais en généralisant l'expérience à toutes les tonalités majeures et mineures. Les profils tonaux que l'on connaît aujourd'hui viennent de cette expérience.

Avec ces profils tonaux, C. L. Krumhansl et E. J. Kessler définissent une mesure de distance entre tonalités en calculant la corrélation entre les profils tonaux de ces deux tonalités. Ils en déduisent une carte en deux dimensions représentant la distance entre tonalités. Elle ressemble très fortement au tableau des relations entre tonalités de Weber (présenté et discuté à la section 5.2.3) et est de forme toroïdale.

C'est dans un article ultérieur [74] que C. L. Krumhansl et M. A. Schmucker mettent au point un premier algorithme estimant la tonalité principale d'une œuvre. Pour ce faire, ils calculent la corrélation entre un histogramme construit selon la distribution des hauteurs d'un certain nombre de notes de la pièce et toutes les tonalités majeures et mineures. La tonalité estimée est celle qui obtient le maximum de corrélation. Le coefficient de corrélation se calcule comme le ratio de la covariance entre les histogrammes sur le produit des écarts-types de chaque histogramme.

Cet algorithme est testé en construisant l'histogramme des 4 premières notes (et les éventuelles notes de l'accord de la quatrième note) sur chacun des 48 préludes du *Clavier Bien tempéré*. L'algorithme choisit la bonne tonalité pour 44 des 48 préludes soit 91.7% de réussite. D'autres tests similaires sont réalisés sur les préludes de Chostakovitch (70.8% de précision), les préludes de Chopin (45.8%) ainsi que les 48 fugues du *Clavier Bien Tempéré* et les 24 fugues de Chostakovitch (91.7% de précision dans les deux cas). Les auteurs imputent les erreurs d'estimation au trop petit nombre de notes pris pour faire l'histogramme de la pièce, l'algorithme étant alors bien trop peu informé pour faire un choix de tonalité réfléchi. Cependant, un plus grand choix de notes peut aussi éloigner l'estimation de la tonalité de la tonalité attendue en considérant dans l'algorithme une note favorisant un autre profil tonal. L'évaluation semble problématique, car 4 notes ne sauraient refléter l'intégralité d'une œuvre musicale.

3.1.3 Monopole des profils tonaux pour l'estimation de tonalité

L'algorithme de Krumhansl-Schmucker suscite depuis sa parution un grand engouement de la part de la communauté. En effet, de nombreux travaux réutilisent cette idée de pondération des notes d'une gamme et de corrélation entre deux profils en changeant les profils tonaux ou la méthode d'évaluation, comme par exemple D. Temperley [129], B. J. Aarden [1], C. Sapp [117], S. Pauws [100], H. Bellmann [12], J. Albrecht et D. Shannahan [5] et N. Napolés López [92]. Les profils tonaux introduits par ces articles sont représentés sur la figure 3.2. L'algorithme est par ailleurs disponible de manière native dans l'environnement music21 dédié à l'analyse musicale computationnelle [35].

Cette méthode inspirera également dans une certaine mesure le modèle présenté dans le chapitre 5.

Temperley. Dans l'article *What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered* paru en 1999 [129], D. Temperley porte un regard critique sur l'algorithme de Krumhansl-Schmuckler et en propose une amélioration suivant trois axes.

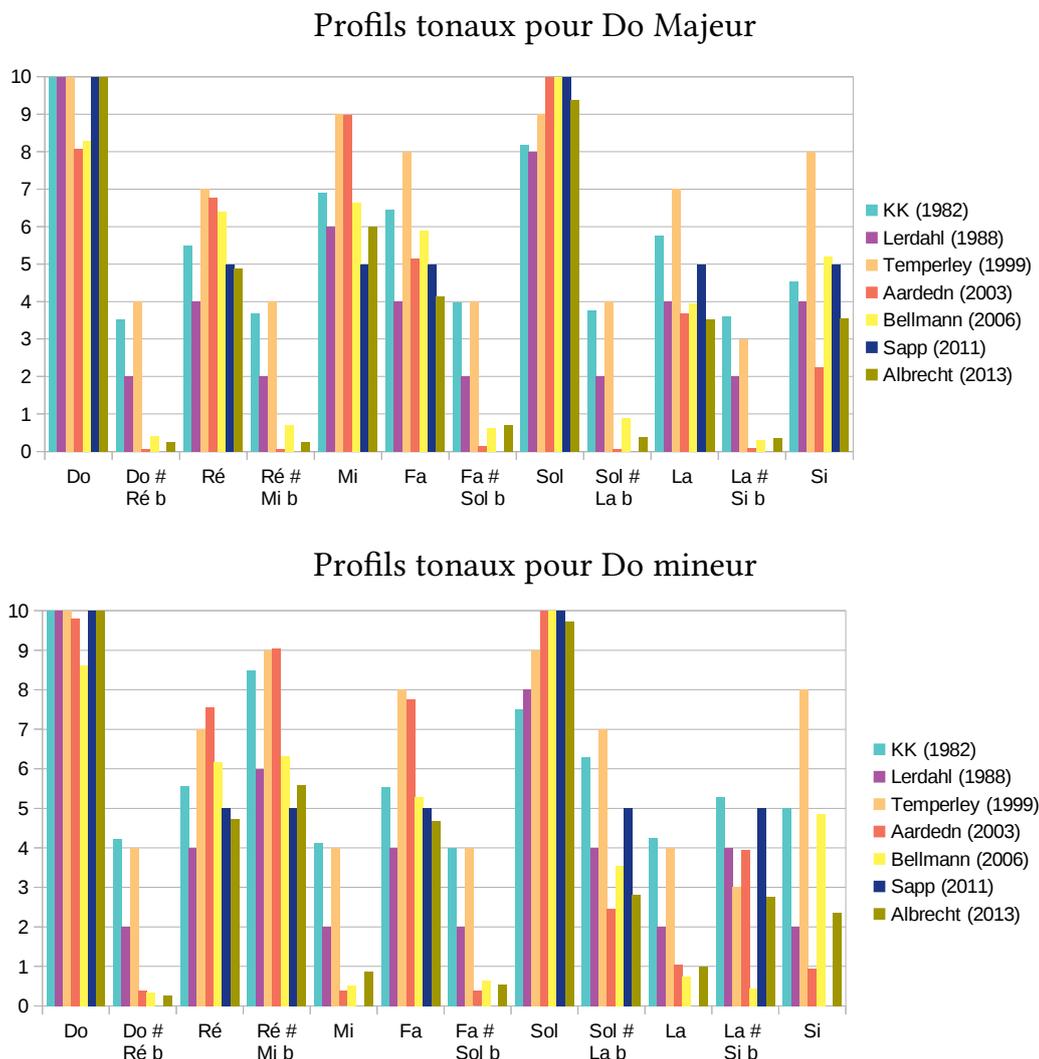


FIGURE 3.2 – Profils tonaux pour Do Majeur et Do mineur introduits par différents auteurs. Les profils tonaux représentés sont ceux de Krumhansl-Kessler (KK,[75]), Lerdahl [78], Temperley [129], Aarden [1], Bellmann [12], Sapp [117] et Albrecht [5]. Cette figure est construite d’après les valeurs trouvées dans les articles respectifs.

Tout d’abord, il suggère de supprimer du calcul de corrélation, la normalisation des vecteurs correspondants au profil tonal et à la pièce. En effet, on ne cherche pas à avoir une valeur de corrélation stricte mais à déterminer quelle est la tonalité la plus adaptée à une œuvre musicale. La normalisation complexifie les calculs. On remplace donc le calcul de corrélation par un simple *produit scalaire* entre les deux vecteurs, bien plus facile à appréhender. La tonalité choisie est alors celle obtenant le score le plus haut.

Ensuite, il propose de changer la pondération des notes dans le profil tonal. Certaines incohérences ont été relevées par D. Temperley, comme par exemple, des valeurs très proches entre $Fa\sharp$ (2.52) et Si (2.88) pour le profil de *Do Majeur* ou la prépondérance du $Si\flat$ sur le Si en

Do mineur. Enfin, la somme des valeurs pondérées est différente entre les profils des tonalités majeures et ceux des tonalités mineures, ce qui amène l'algorithme à choisir plus souvent les tonalités mineures. D. Temperley propose donc des profils tonaux révisés avec des pondérations plus naturelles qui se rapprochent des poids proposés par F. Lerdahl dans son *Tonal Pitch Space* [78]. Ce profil tonal est présent avec d'autres sur la figure 3.2.

Finalement, D. Temperley change la conception de l'histogramme de la pièce. D. Temperley remarque que l'algorithme de Longuet-Higgins présente certaines similarités avec l'approche Krumhansl-Schmuckler. En effet, on peut modéliser une gamme comme étant un profil tonal dont les notes y appartenant ont une valeur de 1 et dont les notes qui n'y appartiennent pas ont une valeur de 0. D. Temperley appelle cela un profil *flat key* par opposition à la pondération apportée par les autres profils tonaux (*weighted key*). De plus, soit chaque note a la même importance et ce indépendamment de sa durée, c'est-à-dire que si une note apparaît dans la partition elle a automatiquement une valeur de 1 dans l'histogramme (*flat input*), soit on pondère son importance relativement à sa durée (*weighted input*). Un *Ré* blanche aura une valeur plus importante qu'un *Mi* croche, par exemple 2 contre 0.5.

D. Temperley estime aussi que cet algorithme est plus efficace pour déterminer la tonalité locale de chaque section d'une œuvre segmentée plutôt que la tonalité globale des œuvres entières. Cette affirmation rend possible la détection des modulations. Cependant, la segmentation étant arbitraire et la modulation possible uniquement entre deux segments, il n'est pas possible de déterminer finement où sont les points de modulation. D. Temperley fixe la durée des segments à une mesure (ou une seconde pour les fichiers audio) dans ses exemples. De plus, comme il ne faut pas changer de tonalité à chaque mesure, D. Temperley ajoute une pénalité si un changement de tonalité est détecté.

D. Temperley teste le modèle sur les 48 sujets de fugues du clavier bien tempéré et 46 extraits du livre de théorie musicale *Tonal Harmony* de S. Kostka et D. Payne [73] divisé en 896 segments. Le modèle estime correctement la tonalité principale pour 42 des 48 fugues (87.5%) ce qui est moins bon que Krumhansl-Schmuckler, mais les conditions de tests sont moins contestables puisque l'algorithme est appliqué à tout le sujet et non seulement à quelques notes. De plus, l'algorithme détecte la modulation pour 2 des 6 sujets modulants. Sur les extraits de *Tonal Harmony*, l'algorithme estime correctement la tonalité de 751 des 896 segments soit une précision de 83.8%.

Pour conclure D. Temperley propose un profil tonal considérant le nom des notes en plus de leur hauteur (*pitch spelling*). Les notes enharmoniques² sont considérées comme des notes différentes, elles ont donc potentiellement des poids différents. Par exemple, pour le profil tonal de *Do Majeur*, la note *Fa♯* a une pondération de 2.0 alors que *Solb* a un poids de 1.5. Ce n'est pas illogique car *Fa♯* est la sensible de l'accord de V/V alors que *Solb* n'a pas de fonction en *Do Majeur*.

Autres pondérations des profils tonaux. C. Sapp [118] propose un profil tonal encore plus simple. Il ne considère que trois pondérations possibles pour les notes (Figure 3.2) :

- Si la note est la tonique ou la dominante de la tonalité, son poids est 2 dans le profil tonal;
- Si la note appartient à la gamme de la tonalité, son poids est 1 (sauf pour *Sib* et *Si* en *Do mineur* qui ont tous les deux une pondération de 0.5);
- Toutes les autres notes ont pour poids 0.

2. Notes ayant des noms différents mais se jouant sur la même touche du clavier, comme *Fa♯* et *Solb* par exemple.

De plus, C. Sapp fait varier la taille des segments utilisés pour estimer la tonalité et propose une visualisation très graphique à l’aide d’un outil appelé *keyscape* [117].

B. J. Aarden [1], H. Bellmann [12] ainsi que J. Albrecht et D. Shanahan [5] proposent également tous les trois de nouvelles pondérations pour les profils tonaux, représentées là encore sur la figure 3.2. Celles-ci sont déterminées par apprentissage statistique, en suivant la fréquence d’apparition des notes dans une tonalité donnée sur leur corpus d’étude, à savoir

- pour Aarden, 1000 chansons extraites aléatoirement du corpus Essen Folksong Collection,
- pour Bellmann, une analyse par H. Budge de 65902 accords issus d’œuvres de 24 compositeurs, relevés et comptabilisés manuellement en 1937,
- pour Albrecht et Shanahan, un ensemble de 492 œuvres composées entre 1700 et 1950.

Ces derniers reconnaissent cependant la limitation de l’approche. En effet, la distribution des notes dans une partition semble être fortement dépendante de l’époque de composition de la partition. L’apprentissage statistique semble améliorer les résultats pour l’estimation des tonalités globales (cf. les résultats pour le corpus des “490 oeuvres de la BD Humdrum” table 3.1). Les profils tonaux de B. J. Aarden et H. Bellmann ont de plus la particularité d’avoir une valeur plus grande pour le degré de dominante que pour celui de tonique.

Au moment de l’écriture de cette thèse, l’étude la plus récente utilisant des profils tonaux pour déterminer les tonalités est le travail de N. Nápoles López et al. [92] en 2019³. Les auteurs utilisent tous les profils tonaux vus plus tôt dans un seul modèle, un modèle de Markov caché dont la distribution des probabilités d’émission de tonalités dépendent des pondérations dans les profils tonaux définis auparavant. Les probabilités de transitions sont quant à elles calculées suivant un tableau de relations entre tonalités similaire aux tables de J. G. Weber ou de A. Schoenberg, décrites dans la section 5.2.3.

I. Quinn et C. W. White démontrent certaines limitations des profils tonaux [104]. En effet, les articles cités précédemment considèrent généralement que toutes les tonalités avec le même mode partagent le même profil. De lors côté, I. Quinn et C. W. White construisent les profils tonaux pour chaque tonalité indépendamment d’après la distribution des hauteurs de notes sur le corpus de Yale YCAC [139] et observent que les profils tonaux sont dépendants de leur tonique et pas seulement de leur mode. Les tonalités ayant beaucoup d’altérations à la clé utilisent plus volontiers toutes les notes chromatiques comparées à leurs homologues plus sobres en altérations qui utilisent plutôt les notes de la gamme associée à la tonalité.

Dans la section 5.6, nous discuterons la performance du modèle de N. Nápoles López confronté à l’approche présentée dans le chapitre 5, ainsi qu’à un modèle neuronal présenté dans la section suivante.

3.1.4 Autres méthodes pour l’estimation de tonalités

D’autres modèles ont été proposés pour estimer les tonalités de façon aussi bien locale que globale. Bien que l’ensemble de cette thèse travaille sur des partitions (*données symboliques*), quelques méthodes appliquées sur des *données audio* sont mentionnées car on observe certaines similitudes dans les approches.

3. Cette thèse présente une collaboration avec N. Nápoles López dont le contenu est disséminé dans la section 4.4 (création d’un corpus) et le chapitre 5 (évaluation de modèles d’estimation de tonalité locale).

Modèles pour les fichiers symboliques. E. Chew [26] définit une mesure de distance d'une séquence de notes à une tonalité à l'aide du *spiral array* [25]. La succession des tonalités est modélisée par une séquence de *frontières* divisant la partition en plusieurs zones de tonalités différentes. Cette approche semble plutôt précise pour trouver le parcours tonal d'une œuvre, cependant sa complexité explose avec l'augmentation de la taille de la pièce. Ce problème peut être résolu partiellement en intégrant de la connaissance musicologique à l'estimation du parcours tonal, pour réduire le nombre de frontières attendues et ajouter la contrainte que la tonalité de la première zone tonale est la même que la tonalité de la dernière zone tonale.

Des modèles récents d'estimation de tonalité locale sur des partitions symboliques utilisent des approches d'apprentissage machine et en particulier des réseaux de neurones. T.-P. Chen et al. [23, 24] ainsi que G. Micchi et al. [86] ont conçu des modèles estimant la tonalité au niveau local, mais aussi une analyse harmonique complète de la partition. Ces modèles distinguent notamment les annotations concernant la tonalité locale et les accords tonicisés. Le modèle de G. Micchi et al. sera utilisé ultérieurement dans le chapitre 5.

Modèles pour les fichiers audios. H. Purwins et al. ont introduit une méthode de tracage de changement de tonalité dans les signaux audio à l'aide de profils *cq*, qu'ils introduisent et qui sont calculés à l'aide du quotient de la fréquence centrale et de la bande passante de chaque filtre, ce qu'on appelle constante *Q* [103]. Concrètement, les profils *cq* sont des profils tonaux adaptés aux fichiers audios. Leur but est d'identifier le centre tonal et ses variations au cours de l'œuvre.

W. Chai et B. Vercoe proposent un modèle basé sur un modèle de Markov caché pour détecter les changements de tonalité [22]. Leur modèle estime d'abord la tonique (ou centre tonal dans leur papier) puis le mode (*Majeur* ou *mineur*).

B. Catteau et al. [21] introduisent un modèle de reconnaissance de gamme et d'accords. Ils font coïncider une gamme majeure à une tonalité majeure et une gamme harmonique mineure à une tonalité mineure, estimant ainsi que leur système est capable de déterminer la tonalité. Il est intéressant de noter que leur modèle se base à la fois sur la version des profil tonaux de D. Temperley [129] et le travail de F. Lerdahl sur les *tonal pitch spaces* [78].

Ö. Izmirlı introduit un modèle qui trouve localement la tonalité sur des fichiers audios. L'originalité de son approche concerne la segmentation de la partition, qui est effectuée par une factorisation matricielle non-négative [68]. Ö. Izmirlı sépare bien les termes *modulation* et *tonicisation* : "Les fonctions secondaires et les tonicisations s'entendent [littéralement] comme de brèves déviations de la tonalité bien établie dans laquelle ils apparaissent – bien que la frontière entre modulation et tonicisation soit poreuse. Une modulation, quant à elle, suscite un glissement de tonique".

H. Papadopoulos et G. Peeters adoptent une approche similaire pour l'estimation locale de tonalité sur des fichiers audio [98]. Leur modèle essaie de segmenter la partition suivant les points de modulation. Ils partent de méthodes d'estimation de tonalité principale et introduisent des dépendances entre tonalités caractérisées notamment par leurs structures harmonique et métrique, afin de les rendre susceptibles de détecter la tonalité de façon locale.

T. Rocher et al. introduisent un modèle qui fournit un couple (accord, tonalité) à chaque frame audio d'un extrait musical. Ce modèle se base sur la méthode de la recherche du *meilleur chemin* dans un graphe [112]. Pour évaluer la distance entre tonalités, ils utilisent les profils tonaux de D. Temperley [129].

L. Mearns et al. utilisent un modèle de Markov caché afin d'estimer les modulations sur des transcriptions audio de chorals de J. S. Bach [84]. Le modèle est entraîné sur des progressions d'accords. Plusieurs distributions de probabilités d'émission sont obtenues d'après deux tables contenant la probabilité qu'un accord existe dans une tonalité donnée (comme par exemple un accord de La mineur dans la tonalité Do Majeur). Ces tables sont construites d'après des travaux de A. Schoenberg et C. L. Krumhansl. Elles ne contiennent pas d'informations à propos des accords de dominante secondaire, ainsi les auteurs ne s'intéressent pas de ce fait aux tonicisations.

En 2014, J. Pauwels et J.-P. Martens présentent un framework probabiliste pour l'estimation simultanée des accords et des tonalités sur des fichiers audio [99]. Bien qu'ils n'utilisent pas du tout les termes de modulation et tonicisation, les auteurs mentionnent l'importance d'intégrer des connaissances musicales au préalable dans les systèmes d'estimation de tonalité locale.

La même année, C. Weiss et al. ont proposé un estimateur de gamme sur les fichiers audio [138]. Ils peuvent déterminer la tonalité courante grâce à la méthode d'analyse des gammes de Z. Gárdonyi [53].

3.2 Perception, modélisation et détection automatisée des cadences

Si la communauté de musicologie computationnelle s'est beaucoup intéressée à l'estimation des tonalités, celle-ci ne s'est que trop peu penchée sur le problème de détection automatisée des cadences, quand bien même celle-ci serait d'une utilité certaine pour la détection de structures et plus particulièrement de la forme musicale.

Une des raisons à cette réticence pourrait être la complexité à modéliser ce phénomène. En effet, les théoriciens de la musique se posent encore aujourd'hui la question de la modélisation des cadences [96]. Dans cet ouvrage, M. Rohrmeier et M. Neuwirth ont suggéré une caractérisation des cadences à l'aide d'une grammaire basée sur les degrés et la ligne de basse [114]. Leurs travaux ont été approfondis par un système modélisant les relations harmoniques entre accords, les règles grammaticales étant inférées pour les harmonies de type jazz [38].

Bien que la notion de cadence soit enseignée dans les livres de formation musicale, les théoriciens de la musique se rejoignent sur la difficulté de définir précisément ce qu'est une cadence. En effet, il y a plusieurs types de cadences, celle-ci se distinguant par la progression harmonique observée mais également par la notion difficilement modélisable de fin de phrase. Une phrase musicale étant très souvent unique⁴, on trouve de nombreuses cadences sous des formes bien différentes dans la littérature musicale [19], rendant de fait complexe leur modélisation informatique.

Une façon de diminuer cette complexité est d'étudier les cadences dans le contexte d'un corpus. C'est ce que font N. J. Martin et J. Pedneault-Deslaurier lors d'une étude des demi-cadences dans les sonates pour piano de Mozart [82] publiée dans le même ouvrage. Cependant, cela limite très fortement le nombre de données d'études et minimise l'impact de l'évolution du style musical du compositeur au cours de sa carrière.

Ainsi, nous ne comptons à ce jour que très peu d'approches pour résoudre le problème de détection des cadences dans les partitions symboliques dont l'une est contemporaine à la

4. Sauf si elle est la reprise, plus ou moins éloignée, d'une phrase entendue précédemment.

méthode proposée dans cette thèse en section 6. Ces approches ont été précédées d'études sur la perception des cadences et en particulier la sensation de fermeture de ces dernières, proposant notamment des modélisations statistiques des cadences. Je commencerai donc par la présentation de ces études avant de passer à la détection automatisée de cadences à proprement parler.

3.2.1 Cadences et perception de fermeture

Si la notion de cadence est principalement associée aux progressions harmoniques qui la caractérisent, elle est aussi indissociable de la notion de fin de phrase. C'est notamment ce que met en exergue W. E. Caplin [19].

D. Sears et al. ont étudié dans plusieurs articles la corrélation entre la perception de cette notion de fermeture et le type de cadence rencontré. Dans une première étude [122], 40 personnes, séparées en 2 groupes suivant leur connaissance de la théorie musicale occidentale (que j'appellerai *musiciens* et *non musiciens*), écoutent un extrait présentant une cadence d'un des 5 types (PAC, IAC, DC, EV et HC) issu des quatuors à cordes de J. Haydn. Ces personnes notent ensuite ces extraits entre 1 et 7, 1 signifiant que le morceau n'est pas terminé alors que 7 signifie que le morceau n'a vraisemblablement pas de raison de continuer. D. Sears et al. demandent également aux participants d'exprimer leur confiance dans leur jugement et leur familiarité avec l'extrait joué par deux autres notes entre 1 et 4. Des résultats de cette expérience, les auteurs tirent les conclusions suivantes :

- Le type de cadence ressenti comme le plus conclusif est PAC suivi d'IAC, HC, DC et EV ce qui confirme l'intuition commune et la théorie de W. Caplin⁵ [17]. Cette théorie s'appelle *Prospective Schema* ;
- Les musiciens ont tendance à écouter la basse pour déterminer si la musique continue alors que les non-musiciens sont plus susceptibles de s'intéresser à la mélodie ;
- La présence de certains éléments dans la musique tels qu'un retard sur le point d'arrivée cadentiel, un trille cadentiel juste avant ce point, la diminution brusque de la densité d'événements sonore au point d'arrivée cadentiel et la durée de l'extrait ont significativement impacté la décision des participants à l'étude ;
- La conclusivité d'un extrait dépend également de sa place dans l'œuvre musicale. Par exemple, une PAC d'un second thème sera plus conclusive qu'une PAC du premier thème.

Cette étude permet de justifier l'utilisation de descripteurs pour identifier les points d'arrivée cadentiels tels que présenté dans le chapitre 6.

Une autre étude par les mêmes auteurs [124] utilise la même approche pour cette fois estimer les attentes des auditeurs suivant le type de cadence. Les résultats obtenus montrent que la cadence la plus attendue est la PAC, suivie de celles se terminant par une harmonie de tonique, que celle-ci soit effective (IAC) ou évitée (DC, EV). La demi-cadence (HC) est la moins attendue, probablement de par son caractère suspensif.

Dans une troisième étude, D. Sears accompagné d'autres collaborateurs utilise le modèle IDyOM [101] sur un corpus constitué d'expositions de quatuors à cordes composés par Haydn afin de montrer que la prédictibilité de la musique augmente à proximité des points d'arrivée de cadence et est minimale juste après ceux-ci [123]. La musique est modélisée suivant les

5. Cela apparaît clairement pour le groupe des musiciens. Pour l'autre groupe, la frontière entre HC, DC et EV est plus floue.

points de vue multiples (*multiple viewpoints*) de D. Conklin et al. [32]. Le modèle comprend 11 points de vue sur la mélodie, 4 sur la ligne mélodique du premier violon, 4 sur la ligne mélodique du violoncelle et 3 sur les accords formés.

Enfin, E. Smit et al. [126] proposent en 2020 une mesure de la perception de différents types de cadences à travers deux valeurs, l'activation psychologique (*arousal*) et la valence, c'est à dire à quel point la cadence est agréable. Par la mise en place d'une expérience ressemblant fortement à celles de Sears et al. évoquées précédemment, ils obtiennent certains résultats, notamment :

- Demi-cadence (HC) et cadence rompue (DC) stimulent plus l'activation car elle brisent les attentes musicales de l'auditeur.
- Les cadences dans une tonalité mineure sont moins activantes et moins attendues que les cadences majeures.
- L'ajout d'une septième à la dominante augmente la valence et l'excitation.
- Plus les hauteurs de notes sont hautes, plus la valence et l'activation sont hautes.

Ainsi, un parcours de ces papiers montre que la cadence est un événement *prédictible* et *attendu* donc *préparé* par d'autres événements musicaux. Ainsi par l'observation de ces préparations, il nous semble possible de *prédire* quels sont les points d'arrivée cadentiels dans les partitions musicales.

Dans la section suivante, je présente d'autres travaux sur la modélisation et/ou la détection des points d'arrivée cadentiels.

3.2.2 Modélisation des cadences pour la prédiction automatisée

P. Kranenburg et al. estiment que la détection des cadences est avant tout un problème de segmentation [133]. Sur un corpus de chansons folk monophoniques néerlandaises (*Meertens Tune Collections*⁶), les auteurs modélisent la partition musicale comme un ensemble de trigrammes de notes⁷ et représentent ces trigrammes comme un ensemble de descripteurs. Ces descripteurs décrivent les hauteurs de notes, le contour, le rythme, le texte, la fermeture telle que la caractérise E. Narmour (modèle Implication-Réalisation [95]), et le contexte. Grâce à un arbre de classification et une validation croisée, les auteurs détectent les cadences dans le corpus et obtiennent une *F*-mesure de 0.80 pour l'expérience la plus fructueuse. A priori, tous les descripteurs semblent utiles pour la détection sauf ceux sur le contour. Les descripteurs les plus significatifs sont ceux sur les silences, sur le texte et ceux basés sur le modèle de E. Narmour. Ainsi, la détection de cadences est un problème *multidimensionnel*.

Dans l'équipe Algomus, de précédents travaux sur les fugues [55] et les formes sonates [15] proposent un système de règles simples consistant à regarder le mouvement de basse pour détecter les cadences, supposées grandement aider à la segmentation de la forme. Les auteurs détectent uniquement les PAC et les rIAC en cherchant les quartes ascendantes et les quintes descendantes à la basse (on suppose alors que cette progression mélodique caractérise une PAC ou une rIAC). Ils ajoutent les contraintes suivantes : l'accord de tonique supposé doit être un accord parfait majeur, mineur ou à tierce retardée (sus4). L'accord de dominante doit être un accord de septième de dominante avec obligation de présence de la sensible. Cette méthode permet de trouver 82% des PAC. La détection des rIAC est plus difficile (66%, et de nombreux faux positifs).

6. 4120 chansons encodées symboliquement sur la base d'enregistrements du début du XX^e siècle

7. Trois notes qui se succèdent. Ils utilisent en particulier deux types de trigrammes, les trigrammes de pitches et les trigrammes de notes qui ignorent les notes répétées.

Enfin, B. Duane et al. mettent en place des clusters harmoniques [43] puis des n -grams mélodiques [42] afin de modéliser la perception musicale de l'auditeur. Cette modélisation permet la détection de cadence (PAC et HC) par l'utilisation dans un premier temps de classificateurs naïfs Bayésiens puis par des modèles de Markov cachés à profil. Bien que les auteurs soient persuadés que cette méthode permette de simuler le comportement de l'auditeur "humain", leurs résultats concernant l'identification de cadence me semblent peu pertinents avec énormément de faux positifs.

3.3 Bilan

Dans ce chapitre, nous avons présenté différents travaux d'informatique musicale concernant l'estimation de tonalités et la détection automatique des cadences.

À propos de l'estimation des tonalités, nous avons vu que les premières approches basées sur des règles, ont été assez vite supplantées par des approches calculant la corrélation entre les notes présentes sur la partition ou l'extrait musical audio et une représentation de la gamme associée à une tonalité ou un profil tonal, qui pondère chaque hauteur de note relativement à son importance dans la gamme. Ces approches sont utilisées aussi bien pour l'estimation des tonalités globales et locales, sur des fichiers audio ou des fichiers symboliques.

Peu d'approches sur les fichiers symboliques existent pour l'estimation de tonalité locale. On compte parmi elles deux modèles utilisant des méthodes d'apprentissage machine et plus particulièrement des réseaux de neurones. La plupart des approches pour l'estimation des tonalités locales sont conçues pour les fichiers audio. Elles ajoutent à leur modélisation pour certaines, une connaissance du contexte temporel à proximité du moment pour lequel la tonalité locale est estimée.

Dans le chapitre 5, nous concevons un modèle d'estimation de tonalité locale sur des fichiers symboliques. Ce modèle calcule des corrélations non pondérées tout en ayant connaissance du contexte temporel, ce qui nous semble indispensable pour proposer une modélisation partielle du système tonal.

À propos de la détection de cadences, peu de travaux existent sur la prédiction des points d'arrivée cadentiels, la plupart des travaux présentés ayant plutôt étudié la perception des cadences. Parmi les trois études existantes, une obtient de bons résultats mais ne porte que sur la musique monophonique. Les deux autres travaillent sur la musique polyphonique mais les résultats peuvent être améliorés par l'utilisation d'autres méthodes. Pour atteindre cet objectif, notre modèle propose une modélisation de la notion de fin de phrase musicale dans le chapitre 6.

Le prochain chapitre présente des jeux de données et des corpus qui ont été établis dans le cadre de cette thèse pour évaluer les modèles.

Chapitre 4

Encodage et annotation de corpus pour la musicologie computationnelle

Contenu

4.1	Corpus annotés de cadences préexistants	43
4.2	Corpus des quatuors à cordes de Mendelssohn : la qualité des données avant tout!	43
4.2.1	Contexte	43
4.2.2	Comparaison des différents stades de corrections	47
4.2.3	Annotations	48
4.3	Corpus des quatuors à cordes de W. A. Mozart : des annotations pour l'analyse musicale	48
4.3.1	Contexte	48
4.3.2	W. A. Mozart et les quatuors à cordes	49
4.3.3	Annotations	50
4.3.4	Utilisation du corpus	55
4.4	Modulations et tonicisations	55
4.4.1	Contexte	55
4.4.2	Références théoriques sélectionnées	55
4.4.3	Encodage et annotations	56
4.4.4	Statistiques	57
4.5	Bilan	59

Lorsqu'on évoque l'apprentissage machine et les données musicales symboliques, un problème majeur est rapidement évoqué : le peu de données d'entrées disponibles, que ce soit pour l'apprentissage¹ ou pour la validation des algorithmes. En particulier, la musique tonale des époques antérieures au XXI^e siècle comporte son lot de spécificités qui rendent très délicat son encodage, à savoir :

- La musique est sur papier et les outils de reconnaissance optique de musique (OMR) ne sont pas assez efficaces et précis pour pouvoir les encoder massivement²
- De nombreux symboles sur les partitions n'ont pas d'équivalent dans les langages informatiques représentant les partitions musicales, en particulier les textes étendus sur plusieurs mesures comme les indications d'expressions, de nuances ou de tempo ;
- Encoder une partition manuellement est long et fastidieux.

Et quand bien même les corpus seraient encodés, les annoter est également une tâche compliquée car longue et potentiellement sujette à interprétation personnelle (voir chapitre 2). Si pour la génération de musique, ces annotations ne sont pas particulièrement nécessaires, elles sont par contre indispensables pour des tâches d'analyse musicale automatisée. Fort heureusement, la communauté d'informatique musicale met déjà à disposition certains corpus utiles à diverses tâches telles que la génération, la séparation des sources ou la caractérisation du sentiment dans la musique. Ces corpus sont disponibles pour la plupart dans un dossier de corpus ISMIR à l'adresse <https://github.com/ismir/mir-datasets>.

Durant ma thèse, j'ai utilisé des *corpus déjà annotés* sur les cadences (Section 4.1). J'ai surtout contribué à l'élargissement de la base de données des partitions encodées numériquement, à celle des annotations musicologiques ainsi qu'à une réflexion sur cette préparation des corpus en *établissant trois corpus* ayant chacun leur spécificité que je présente tout au long de ce chapitre ainsi que dans la table 4.1.

Le premier corpus est celui des *quatuors à cordes de F. Mendelssohn*. Il a été présenté lors de la conférence ISMIR2020 [39]. Pour ce corpus, l'équipe *Computational Tonal Studies*, à laquelle j'appartiens, est partie de la partition papier et a utilisé un logiciel d'OMR pour produire une première version encodée du corpus. Nous avons particulièrement étudié comment la qualité de ce corpus a évolué au cours de corrections successives (section 4.2).

Le second corpus est celui des *quatuors à cordes de W. A. Mozart*. Il est présenté à la section 4.3. Pour ce corpus, les fichiers existaient déjà sous forme numérique (fichiers .krn) et nous y avons ajouté les annotations de modulations, de forme et de cadences. Ce corpus a été central à la fois pour les travaux de cette thèse, détaillés dans les chapitres suivants, et pour les publications de l'équipe [7, 46, 45].

Le troisième corpus est un corpus spécialisé pour les **modulations** constitués de *petits exemples extraits de livres de théorie musicale*. Il a été conçu spécifiquement pour la détection de modulations qui sera discutée dans le chapitre 5. Ce corpus est une collaboration entre Néstor Nápoles López (Université McGill, Montréal) et moi-même. Il a été présenté à l'occasion de la conférence DLfM2020 [93] (section 4.4).

1. Et c'est d'autant plus problématique pour les réseaux de neurones nécessitant de gros volumes de données pour être pleinement efficaces.

2. Il existe cependant quelques études ayant utilisé l'OMR pour soit encoder les corpus, comme celle de J. Cummings et al. pour des corpus de la renaissance [34], ou directement récupérer des annotations ; écrites sur les partitions comme celle de N. Condit-Schultz et al. sur des chorals de Bach [31].

Corpus	Taille	Enc.	Vérif.	Ann.	Articles
Mendelssohn	28 mouvements	X	X		[39]
Mozart	61 mouvements		(X)	X	[7, 46, 45]
Modulations	201 exemples	X	(X)	X	[93]

TABLE 4.1 – Présentation des corpus produits pendant la thèse. Pour chaque corpus, un X est inscrit dans les cases correspondant aux actions effectuées, à savoir l’encodage (Enc.), la vérification/correction (Vérif.) et la modélisation et l’annotation (Ann.). On note un (X) lorsque la vérification était partielle et rapide. On précise enfin dans quels articles les corpus ont été présentés ou utilisés.

Corpus	haydn-quartets	bach-wtc-i
œuvres	Quatuors à cordes, Haydn [123]	Fugues du Clavier bien tempéré, Bach [55]
Pièces	42 expositions	24 fugues
Voix	4	de 2 à 5
Temps	7173	4739
PAC (dont finales)	99 (21)	63(23)
rIAC	(8)	24
HC	70	(5)

TABLE 4.2 – Deux corpus avec des annotations manuelles de cadences existant avant cette étude. On trouve des cadences sur environ 2% des temps. On ne considère que les types de cadences les plus représentés.

4.1 Corpus annotés de cadences préexistants

Pour la première étude du chapitre 6, nous avons utilisé deux corpus existants constitués de cadences (Table 4.2) :

Le corpus `bach-wtc-i` contient les 24 fugues du premier livre du *Clavier bien tempéré* composé par J.-S. Bach. Les annotations de cadences sont issues d’une étude précédente effectuée par l’équipe Algomus [55]. Bien que ces annotations aient été faites dans le cadre d’une détection automatisée des éléments saillants d’une fugue et plus généralement dans l’analyse de la forme, elles sont pertinentes pour la détection des cadences.

Le corpus `haydn-quartets` est constitués de 42 expositions de mouvements en forme sonate des *quatuors à cordes* de Haydn. Les annotations pour ce corpus proviennent d’un article écrit par David Sears et son équipe [123].

Peu de corpus existent avec des cadences annotées, c’est pourquoi nous avons utilisé ces corpus pour nos expériences.

4.2 Corpus des quatuors à cordes de Mendelssohn : la qualité des données avant tout!

4.2.1 Contexte

Ce corpus est l’initiative de l’équipe *Computational Tonal Studies (CTS)* composée de chercheurs s’intéressant à la musicologie, la théorie de la musique, la cognition musicale et la musicologie computationnelle mais surtout à la musique tonale symbolique. Cette jeune équipe a

été fondée en mai 2019 et est composée de 8 chercheurs affiliés à l’université McGill à Montréal et de moi-même affilié à l’Université de Lille.

Nous voulions expressément un corpus tonal, complet, relativement court, qui soit le travail d’un seul compositeur et qui soit de la période classique. En effet, cette période, en plus d’avoir été remise au goût du jour par certains théoriciens, est celle qui semble présenter la musique la plus caractéristique du système tonal. De plus, ce corpus étant destiné à tester, améliorer et concevoir des algorithmes d’annotation automatisée tels que ceux présentés dans la suite de cette thèse, nous l’avons choisi de façon à ce que le travail informatique préliminaire à l’application d’un algorithme de MIR soit le plus réduit possible. C’est pourquoi nous avons réduit notre choix à des corpus de quatuors à cordes. Les quatuors à cordes sont composés de quatre voix indépendantes évitant ainsi le travail préalable de séparation des voix³. De plus, la conduite des voix n’est pas limitée par une contrainte technique comme la taille des mains de l’interprète lorsque l’instrument joué est le piano. Il est donc plus facile d’identifier le rôle de chaque voix (mélodie, accompagnement...) dans un quatuor à cordes.

Les quatuors à cordes des compositeurs classiques ont été vite mis de côté car :

- Les quatuors d’Haydn sont beaucoup trop nombreux ;
- Les quatuors de Mozart étaient en cours de traitement (voir la section suivante 4.3) ;
- Les quatuors de Beethoven étaient déjà annotés [97].

C’est pourquoi notre choix s’est porté sur les quatuors de Mendelssohn. En effet, bien que Mendelssohn soit plutôt un compositeur du début de la période romantique, il apparaît comme le successeur naturel de Haydn, Mozart et Beethoven dans l’écriture stylistique de ses quatuors et ses œuvres gardent encore de nombreuses caractéristiques du style classique telle qu’une harmonie, une organisation formelle et une prosodie très claire. On peut aussi entendre ces caractéristiques dans les derniers quatuors de Beethoven [131].

Les quatuors à cordes composés par Mendelssohn sont présentés dans la table 4.3. Mendelssohn a composé 6 quatuors en 4 mouvements ainsi que 4 pièces indépendantes pour Quatuors à cordes, réunis en un opus (Op. 81) par l’éditeur à la mort de Mendelssohn.⁴ Le dernier quatuor Op. 80 est également publié de façon posthume.

L’édition *Breitkopf und Härtel* de 1875 est utilisée pour l’encodage. Elle est disponible pour tous les quatuors sur *IMSLP*. En partant des partitions sous forme de fichiers .pdf, la première étape fut de générer des fichiers exploitables par les algorithmes, en l’occurrence, des fichiers .musicxml⁵.

Nous avons utilisé le logiciel commercial de *reconnaissance optique de musique (OMR) Photo-score* pour générer une première version C0 des fichiers .musicxml. La précision de ce type de logiciel étant plutôt faible, nous nous attendions à ce que les fichiers obtenus contiennent beaucoup d’erreurs. Il nous a cependant semblé plus efficace d’utiliser cette méthode pour générer une version de base de ces fichiers plutôt que de les recopier “à la main” à l’aide d’un logiciel d’édition musicale.

Ces fichiers C0 ont ensuite été répartis entre les différents membres de l’équipe CTS et corrigés. Ces fichiers corrigés sont en version C1. Pour établir les fichiers C1, des règles concernant certains symboles ont été établies afin d’obtenir une certaine cohérence. Nous avons notam-

3. Bien que ce dernier soit automatisable [59, 60]

4. Il semblerait cependant que les deux premières pièces de cet ensemble auraient été 2 mouvements d’un septième quatuor.

5. MusicXML fut choisi pour deux raisons, non seulement il est possible d’y représenter les données musicales très précisément, mais aussi beaucoup de logiciels acceptent ce format, que ce soit en entrée ou en sortie, ce qui augmente la réutilisabilité du corpus.

Id.	Opus	Composition	Tonalité
1	Op. 12	1829	<i>Mib Majeur</i>
2	Op. 13	1827	<i>La Majeur</i>
3	Op. 44/1	1838	<i>Ré Majeur</i>
4	Op. 44/2	1837	<i>Mi mineur</i>
5	Op. 44/3	1838	<i>Mib Majeur</i>
6	Op. 80	1847	<i>Fa Majeur</i>
FP	(iv)	1827	<i>Mib Majeur</i>
	(iii)	1843	<i>Mi mineur</i>
	(i)	1847	<i>Mi Majeur</i>
	(ii)	1847	<i>La mineur</i>

TABLE 4.3 – Inventaire des pièces pour quatuors à cordes publiées de Mendelssohn. Pour les six quatuors, la tonalité est celle du quatuor et correspond à la tonalité du premier et du dernier mouvement a minima. L’opus 81 réunit quatre pièces indépendantes (Four Pieces, FP).

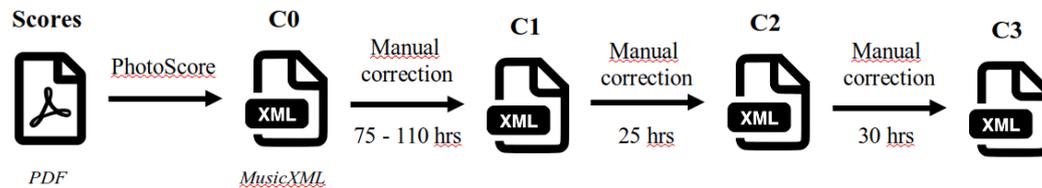


FIGURE 4.1 – Processus d’encodage et de correction du corpus Mendelssohn avec une estimation large du temps passé pour chaque étape.

ment décidé de nous concentrer sur l’exactitude des hauteurs de notes et des rythmes, ainsi que de supprimer certains symboles tel que les textes de la partition, ceux-ci étant une source d’erreur potentielle de par leur difficulté d’interprétation par les algorithmes.

Nous avons ensuite procédé à une vérification des corrections par un autre membre de l’équipe que celui chargé de produire la version C1 des fichiers. Ces fichiers vérifiés sont en version C2. Afin d’avoir une plus grande cohérence, il a aussi été décidé de supprimer plus généralement les symboles musicaux dont la correction prenait beaucoup de temps et qui ne semblaient pas être intéressants pour les algorithmes de musicologie computationnelle comme les nuances et tout particulièrement les *crescendos* et *decrescendos*.

Enfin, la dernière version C3 de ces fichiers correspond à une ultime vérification de ces fichiers par une unique personne ainsi qu’à une standardisation de la mise en page et des métadonnées.

Le processus d’encodage est schématisé à la figure 4.1.

La différence quantitative mesurée entre les fichiers issus des différentes phases de correction est présente à la table 4.4. Nous avons utilisé la librairie Python `difflib` pour générer ces résultats. Ces résultats auraient pu être générés via d’autres outils plus performants comme celui de F. Foscarin et al. [51] mais nous n’en avons malheureusement pas à disposition au moment de la publication.

À la fin de cette phase d’encodage, nous avons exactement 100 fichiers `.musicxml` et `.mscz` (format musescore, le logiciel d’édition utilisé pour les corrections), un pour chaque mouvement et chaque étape de correction sur tous les quatuors, sauf pour les 4 fichiers corres-

The image displays two systems of musical notation for measures 60 to 62 of the 4th movement of Mendelssohn's 5th string quartet. Each system consists of four staves: Violin I (top), Violin II, Viola, and Cello/Double Bass (bottom). The key signature is three flats (B-flat, E-flat, A-flat) and the time signature is common time (C). The upper system (C0) shows the original OMR output, with some notes and accidentals appearing slightly off or with artifacts. The lower system (C3) shows the score after three rounds of corrections, where the notation is cleaner and more accurate, reflecting the improvements made by the correction system.

FIGURE 4.2 – 4^e mouvement du 5^e quatuor à cordes de Mendelssohn Op. 44, No iii, mesures 60 à 62. Le système supérieur représente la sortie d'OMR (C0). Le système inférieur représente la partition après trois tours de corrections (C3).

Catégories comparées	Différence médiane [min, max]
C0 et C1	10.0% [2.8%, 21.8%]
C1 et C2	1.3% [0.0%, 7.3%]
C2 et C3	0.2% [0%, 1.3%]

TABLE 4.4 – Différence en pourcentage entre chaque paire de catégorie de fichiers. Les résultats présentent la médiane calculée sur les 24 mouvements, avec les valeurs maximales et minimales entre crochets.

pendant au quatuor opus 80. En effet, ce quatuor était déjà encodé sous la forme d'un fichier .mscz, nous l'avons donc juste séparé en quatre mouvements puis nous lui avons appliqué le processus de standardisation. Puisque ce fichier n'a pas été produit par nos soins, nous n'en intégrons pas les résultats dans la table 4.4.

Fichiers comparés	Différence d'annotations médiane(%) [min(%), max(%)]
C0 et C1	46.8% [9.9%, 71.1%]
C1 et C2	0.4% [0.0%, 9.8%]
C2 et C3	0.0% [0.0%, 3.0%]

TABLE 4.5 – Proportion de différence des annotations concernant l'estimation locale de tonalité entre deux types de fichiers. On calcule la médiane sur les 24 mouvements ainsi que les valeurs minimales et maximales entre crochets.

Fichiers comparés	Différence de PACs détectées	Nouvelles PACs	PACs perdues
C0 et C1	154.5% de 22 à 56	177.3% 39	22.7% 5
C1 et C2	-3.6% de 56 à 54	1.8% 1	5.4% 3
C2 et C3	1.9% de 54 à 55	3.7% 2	1.9% 1

TABLE 4.6 – Changement dans le nombre de PAC détectées. On recense les cadences nouvellement détectées, ainsi que les cadences qui ne le sont plus. Comme les cadences détectées par l'algorithme sont dénombrables, on affiche à la fois le nombre de cadences et la proportion de changement.

4.2.2 Comparaison des différents stades de corrections

Nous avons appliqué 4 algorithmes de MIR sur ces différents fichiers. Cela a permis de valider la cohérence du corpus et de mesurer l'importance de la correction des fichiers en sortie d'OMR. Les algorithmes appliqués sur le corpus sont :

- un algorithme d'étiquetage d'accords [70],
- un algorithme d'estimation locale des tonalités [92] présenté dans le chapitre 3,
- un algorithme de détection de motifs [30],
- l'algorithme de détection des cadences parfaites [14] présenté en détail dans le chapitre 6.

Comme ce corpus n'intègre pas, pour l'instant, d'annotations de référence, nous ne pouvons mesurer que la différence des annotations générées par les algorithmes et nous n'évaluons pas leur qualité. Les résultats pour l'estimation des tonalités et la détection de cadences sont présentés dans les tables 4.5 et 4.6. L'intégralité des résultats est disponible dans l'article.

On remarque notamment que l'essentiel des différences entre les annotations générées apparaissent entre C0 et C1. Bien que les partitions soient similaires à 90% (cf. table 4.4), les 10% de différence génèrent énormément de différences en sortie des algorithmes (au moins la moitié des annotations changent) et ce quelle que soit la tâche considérée, même si l'estimation de tonalité est un peu moins impactée. Beaucoup de hauteurs de notes ont été modifiées dans cette étape de correction, ce qui explique cette si grande différence. Les corrections suivantes ne semblent avoir que peu d'impact sur les algorithmes, ce qui était attendu puisqu'elles ne modifient elles-mêmes que très peu la partition.

Ainsi, s'il semble nécessaire de corriger la partition générée par l'OMR sous peine de perdre beaucoup de résultats générés par les algorithmes. Cette correction n'a pas besoin d'être minutieuse si le but est juste d'obtenir des annotations automatisées pertinentes.

Cela ouvre également des pistes quant à l'utilisation conjointe de l'OMR et d'algorithmes de CMÀ comparée à une annotation de référence pour la correction automatisée de fichiers en sortie d'OMR.

4.2.3 Annotations

À l'heure actuelle, l'annotation du corpus n'existe pas. Il s'agit de la prochaine étape pour ce corpus. Nous prévoyons des annotations de référence concernant la forme, les cadences, la tonalité, l'analyse harmonique et l'étiquetage d'accords. Nous espérons aussi afficher des annotations d'analyse des fonctions formelles telles celles décrites par W. Caplin [17], et donc identifier les *periods*, les *sentences* et autres éléments de l'analyse fonctionnelle.

4.3 Corpus des quatuors à cordes de W. A. Mozart : des annotations pour l'analyse musicale

4.3.1 Contexte

Lorsque j'ai démarré ma thèse, mes encadrants m'ont fortement encouragé à établir des annotations de référence sur un corpus afin d'avoir un nombre de données intéressant pour l'entraînement d'algorithmes d'apprentissage machine et la validation d'algorithmes en tout genre. Ainsi, l'analyse complète du corpus des quatuors à cordes de W. A. Mozart a été le fil rouge qui a guidé ma thèse.

Le choix de ce corpus fut assez naturel. Il répond à la même logique que le choix des quatuors de F. Mendelssohn. W. A. Mozart est un compositeur de la période classique et son langage harmonique est clair mais non dénué de sensibilité et de surprises, les harmonies sont bien définies mais souvent subtiles, la prosodie est tout aussi claire, tantôt comique, tantôt dramatique. Rappelons-nous aussi que la forme sonate n'a pas encore été théorisée quand W. A. Mozart écrit ses quatuors. Avec J. Haydn, il est le compositeur qui fixe le cadre de cette forme, n'hésitant cependant pas à le faire exploser quand celui-ci restreint ses ambitions composition-

nelles. Pour l’anecdote, lorsqu’il m’a fallu extraire des exemples de ces quatuors pour illustrer certains concepts précis pour les différents articles, comme une demi-cadence ou une césure médiane, il m’a toujours été très compliqué de trouver l’exemple standard tel qu’on le présente dans les livres de théorie musicale. Dans les exemples retenus, il y avait toujours une note, une nuance, un silence inopportun qui brisait les attentes de l’auditeur et de fait les miennes.

L’autre avantage de ce corpus est qu’une grande partie des mouvements qui le composent est déjà encodée numériquement sous forme de fichiers Humdrum `.kern` [67]. Parmi les 86 mouvements composés par Mozart, 72 sont encodés. Ces fichiers ont été récupérés via <http://github.com/musedata/humdrum-mozart-quartets>. Ils étaient initialement disponibles sur <http://kern.humdrum.org> et encodés par E. Correia, Jr. et F. Bennio.

L’essentiel du travail a donc été de produire une annotation de référence établissant pour chaque mouvement la forme de ce dernier, les tonalités ainsi que les cadences. Cette étude a été l’occasion pour moi de plonger dans l’écriture musicale de W. A. Mozart et de suivre son évolution à travers ses quatuors.

4.3.2 W. A. Mozart et les quatuors à cordes

W. A. Mozart a écrit 23 quatuors à cordes au cours de sa vie. Leur écriture donne un bon aperçu de son évolution stylistique en tant que compositeur.

Son premier quatuor (K 80) a été composé en 1770 dans une taverne sur la route de Milan à Bologne. W. A. Mozart était alors âgé de 14 ans⁶. La structure de ce quatuor originellement en 3 mouvements (Adagio-Allegro-Menuetto) ainsi que son style sont inspirés par les quatuors de G. B. Sammartini sur le modèle de la sonate en trio. Il marque un tournant dans le processus créatif de Mozart [71]. Un quatrième mouvement est ajouté en 1773 ou 1774 alors que Mozart et son père se trouvent à Vienne.

Les six quatuors suivants (K 155 à K 160), connus comme les *quatuors milanais*, ont été composés en 1772 et 1773 durant le troisième voyage de W. A. Mozart en Italie du nord. Le premier quatuor a été composé à Bolzano, les cinq autres à Milan. Ces quatuors assoient le style musical mozartien même si celui-ci garde les sigmates du *style galant* propre à l’opéra italien [102].

Les six quatuors K 168 à K173 sont surnommés les *quatuors viennois* car composés à Vienne à la fin de l’année 1773. L’influence des quatuors à cordes opus 17 et 20 de J. Haydn se fait clairement ressentir dans ces compositions de W. A. Mozart, bien que ce dernier semble avoir quelques difficultés à absorber toutes les subtilités et les innovations de la musique de J. Haydn [87]. Cependant, W. A. Mozart y montre une grande habileté dans l’écriture contrapuntique. En effet, deux des mouvements finaux de ces quatuors sont des fugues.

Les six quatuors dédiés à J. Haydn (K 387, K 421, K 428, K 458, K 464 et K 465) comptent sans aucun doute comme les pages les plus fines et célèbres du compositeur dans le style de quatuor et probablement dans son œuvre intégrale. Ils sont “le fruit d’un travail long et laborieux” selon les propos de W. A. Mozart dans la dédicace attachée à la publication et furent composés entre 1782 et 1785. Plus précisément, K 387 fut composé en décembre 1782, K 421 en juin 1783⁷, K 428 en juillet 1783, K 458 (“*La chasse*”) à la fin de 1784, K 464 et K 465 (“*Les disso-*

6. Mozart est né en janvier 1756 et est mort en décembre 1791.

7. Contemporain de la naissance du premier fils de Mozart, il est intéressant de noter que la tonalité de ce quatuor est *Ré mineur*, une tonalité lourde de sens pour Mozart que l’on rencontrera notamment dans le *Concerto pour piano no 20*, l’ouverture de *Don Giovanni* et le *Requiem*

nances”) en janvier 1785. Ce dernier quatuor, surnommé d’après son étrange introduction⁸ fut le sujet de nombreuses questions au moment de sa création. L’éditeur *Arturia* a même renvoyé le manuscrit à Mozart prétextant qu’il était “plein d’erreurs” et “injouable”. Ce n’est pas l’opinion d’Haydn qui tint le 12 février 1785 (soit juste après la création des trois derniers quatuors du cycle) auprès de Leopold Mozart, le père de Wolfgang, l’éloge suivant (traduit depuis [91], voir aussi [136]) :

Je vous le dis devant Dieu, en honnête homme, votre fils est le plus grand compositeur que je connaisse, en personne ou de nom, il a du goût, et en outre la plus grande science de la composition.

Un quatuor isolé K 499 fut publié en août 1786. Il est nommé “Hoffmeister” d’après la dédicace de Mozart à son ami, compositeur et éditeur F. A. Hoffmeister. C’est l’une des pièces les moins connues de Mozart. Les musicologues semblent partagés quant à sa qualité. A. Poirier dans [102] trouve le premier mouvement “absolument merveilleux” alors que H. King dans [71] y voit une “page étrange [à laquelle] il manque la spontanéité et l’émotion caractéristiques des quatuors dédiés à Haydn, elle ne leur en est en aucun cas inférieure.”

Les trois derniers quatuors (K 575 en juin 1789, K 589 en mai 1790 et K 590 en juin 1790) sont les trois premiers quatuors d’une commande de six quatuors de la part du roi Friedrich Wilhelm II de Prusse et sont surnommés à ce titre les *quatuors prussiens*. La mort de Mozart en 1791 laissera ce cycle inachevé. Bien que l’écriture des partitions fut “pénible” selon les mots de Mozart lui-même, la musique révèle une virtuosité prononcée et une beauté certaine dans la sonorité des cordes, particulièrement dans la partie de violoncelle, Friedrich Willem II étant lui-même un violoncelliste talentueux. Mozart met en lumière l’unité thématique ainsi qu’une “utilisation brillante et vigoureuse du fugato dans les mouvements finaux” [71]. Ces ultimes quatuors furent publiés après la mort de compositeur.

4.3.3 Annotations

Sur les 86 mouvements que compte l’ensemble des quatuors à cordes de Mozart, 61 ont fait l’objet d’une annotation de référence durant la thèse (Table 4.7). Ces annotations de référence ont été d’abord posées sur partition papier puis encodées numériquement via *Dezrann* [54], l’application web développée par l’équipe Algomus, et sont représentées sous un format json (fichier `.dez`) tout en rendant possible différents exports sous forme de texte simple. Un exemple de ce fichier *Dezrann* est donné à la figure 4.3.

Pour tous les mouvements étudiés, les annotations portent sur la forme musicale du mouvement, la tonalité et les cadences. Pour certains des mouvements, je fournis également l’analyse harmonique mais celle-ci n’étant pas nécessaire pour nos études et très longue à inscrire numériquement, je ne l’ai pas systématiquement encodée⁹.

Ces annotations ont été revues et corrigées (et discutées en cas de désaccord) par d’autres membres de l’équipe.

Annotation des tonalités Afin de garantir l’unicité d’un label concernant les tonalités à tout moment de la partition, nous considérons qu’une seule tonalité est possible à la fois. Nous excluons de fait l’ambivalence des accords pivots et nous considérons qu’ils ne sont que dans

8. Elle ferait référence à l’initiation franc-maçonnique [87]

9. Par contre, elle est effectivement écrite sur papier car indispensable pour déterminer la tonalité et les points d’arrivée cadentiels.

Oeuvre	Tonalité Globale	Forme	Composition	Tonalité quatuor	Type MC
K80	Mvt 1	Sol Majeur	Mars 1770	Sol Majeur	1
	Mvt 2	Sol Majeur			2
	Mvt 3	Sol Majeur			NA
	Mvt 4	Sol Majeur			NA
K155	Mvt 1	Ré Majeur	Automne 1772	Ré Majeur	3
	Mvt 2	La Majeur			2
	Mvt 3	Ré Majeur			NA
K156	Mvt 1	Sol Majeur	Fin 1772 / Début 1773	Sol Majeur	1
	Mvt 2	Mi mineur			2
	Mvt 3	Sol Majeur			NA
K157	Mvt 1	Do Majeur	Début 1773	Do Majeur	1
	Mvt 2	Do mineur			4
	Mvt 3	Do Majeur			NA
K158	Mvt 1	Fa Majeur	Début 1773	Fa Majeur	1
	Mvt 2	La mineur			2
	Mvt 3	Fa Majeur			NA
K159	Mvt 1	Si bémol Majeur	Début 1773	Si bémol Majeur	1
	Mvt 2	Sol mineur			3
	Mvt 3	Si bémol Majeur			NA
K160	Mvt 1	Mi bémol Majeur	Début 1773	Mi bémol Majeur	1
	Mvt 2	La bémol Majeur			2
	Mvt 3	Mi bémol Majeur			1
K168	Mvt 1	Fa Majeur	Août 1773	Fa Majeur	2
	Mvt 2	Fa mineur			1
	Mvt 3	Fa Majeur			NA
	Mvt 4	Fa Majeur			NA
K169	Mvt 1	La Majeur	Août 1773	La Majeur	4
	Mvt 2	Ré Majeur			3*
	Mvt 3	La Majeur			NA
	Mvt 4	La Majeur			NA
K170	Mvt 1	Do Majeur	Août 1773	Do Majeur	NA
	Mvt 2	Do Majeur			NA
	Mvt 3	Sol Majeur			2
	Mvt 4	Do Majeur			NA
K171	Mvt 1	Mi bémol Majeur	Août 1773	Mi bémol Majeur	1
	Mvt 2	Mi bémol Majeur			NA
	Mvt 3	Do mineur			1
	Mvt 4	Mi bémol Majeur			1
K172	Mvt 1	Si bémol Majeur	Août 1773	Si bémol Majeur	1
	Mvt 2	Mi bémol Majeur			1
	Mvt 3	Si bémol Majeur			NA
	Mvt 4	Si bémol Majeur			2
K173	Mvt 1	Ré mineur	Août 1773	Ré mineur	1
	Mvt 2	Ré Majeur			NA
	Mvt 3	Ré mineur			NA
	Mvt 4	Ré mineur			NA
K387	Mvt 1	Sol Majeur	Décembre 1782	Sol Majeur	1
	Mvt 2	Sol Majeur			1
	Mvt 3	Do Majeur			1
	Mvt 4	Sol Majeur			1
K421	Mvt 1	Ré mineur	Juin 1783	Ré mineur	3
K428	Mvt 1	Mi bémol Majeur	1783	Mi bémol Majeur	3
	Mvt 2	La bémol Majeur			2
K458	Mvt 1	Si bémol Majeur	Novembre 1784	Si bémol Majeur	3
K465	Mvt 1	Do Majeur	Janvier 1785	Do Majeur	1
	Mvt 4	Do Majeur			1
K499	Mvt 1	Ré majeur	Août 1786	Ré Majeur	NA
	Mvt 3	Sol Majeur			1
K589	Mvt 1	Si bémol Majeur	Mai 1786	Si bémol Majeur	3
K590	Mvt 1	Fa Majeur	Juin 1786	Fa Majeur	1

TABLE 4.7 – Les 61 mouvements du corpus des quatuors de Mozart ont été analysés avec plus de 2000 annotations. Ces annotations de référence sont disponibles en licence open-source à l'adresse <http://www.algomus.fr/data>.

6/4	7/+	5	5	#	6/4	7/+	5
V (I)	V	I:PAC	II	V	V (I)	V	I

```

"meta": {
  "title": "String Quartet #02 in D major, K 155, mvt. 1",
  "name": "algonus",
  "date": "2019-09-02T15:16:11.154Z",
  "producer": "Dezrann c6c80150",
  "cite": "P. Allegraud et al., Learning sonata form structure on Mozart's string quartets",
  "curator": "Laurent Feisthauer",
  "source": "http://www.algonus.fr/data"
},
"labels": [
  { "type": "Tonality", "start": 0, "duration": 46, "line": "bot.1", "tag": "D"},
  { "type": "Tonality", "start": 48, "duration": 162, "line": "bot.1", "tag": "A"},
  { "type": "Tonality", "start": 306, "duration": 7, "line": "bot.1", "tag": "Bb"},
  { "type": "Cadence", "start": 44, "line": "all", "tag": "I:PAC"},
  { "type": "Cadence", "start": 108, "line": "all", "tag": "V:PAC"},
  { "type": "Cadence", "start": 200, "line": "all", "tag": "V:PAC EEC"},
  { "type": "Cadence", "start": 456, "line": "all", "tag": "I:PAC ESC"},
  { "type": "Cadence", "start": 108, "line": "all", "tag": "V:PAC MC"},
  { "type": "Structure", "start": 0, "duration": 210, "line": "top.1", "tag": "EXPOSITION"},
  { "type": "Structure", "start": 212, "duration": 70, "line": "top.1", "tag": "DEVELOPMENT"},

```

FIGURE 4.3 – 1^{er} mouvement du quatuor à cordes no 2 en Ré majeur de W. A. Mozart, K 155, mesures 10 à 16 . Capture d'écran issue de Dezrann (www.dezrann.net). Dezrann permet d'inscrire les annotations directement sur la partition numérique. Ici sont annotées la forme (trois lignes en jaune au dessus du texte musical, la première correspond à la grande partie (ici l'exposition), les deux suivantes aux sections découpant ces parties (fin de P, début de TR) pour mon annotation et l'annotation de Flothuis [49]), la tonalité (en mauve en dessous du texte musical), les cadences (barre verticale violette au premier temps de la mesure 12 indiquant le point d'arrivée cadentiel d'une PAC) ainsi que l'analyse harmonique (en dessous de la partition) et certaines notes étrangères (en vert sur les têtes de notes). En-dessous de la capture d'écran se trouve un extrait du fichier .dez correspondant à l'analyse de l'œuvre. Chaque label défini sur la partition est encodé numériquement sur une ligne dans le fichier. On montre deux correspondances pour la cadence et pour la nouvelle tonalité.

une seule tonalité¹⁰. Pour identifier les modulations et déterminer les tonalités, nous cherchons plusieurs indicateurs. Le premier est la présence d'une note qui ne ferait pas partie de la gamme de la tonalité actuelle. Celle-ci peut soit être une note étrangère soit une note appartenant à une autre gamme, celle de la nouvelle tonalité. Nous regardons alors l'accord qui contient cette note altérée ainsi que ceux l'entourant et nous essayons d'y adjoindre une analyse harmonique plausible dans une nouvelle tonalité (de type II, V, I, etc.). Dans la plupart des cas, la note altérée appartient à un accord de V de la nouvelle tonalité et s'enchaîne à un I. Le second est

10. Nous nous autorisons cependant une exception quand les accords sont sur pédale. Nous considérons alors le début de la pédale comme étant dans la tonalité de départ et la fin comme étant dans la tonalité d'arrivée

la présence d'une cadence. En effet, après une cadence, on commence une nouvelle phrase voire une nouvelle section et cette phrase peut fort bien être dans une nouvelle tonalité sans que celle-ci ait été préparée en amont, puisque l'événement après une cadence est le moins prédictible de la partition [123]. Nous regardons donc les progressions harmoniques et nous cherchons à identifier quelques schémas qui donnent des indices sur la tonalité [57]. De plus, certaines sections ont des tonalités obligées, on s'attend donc à moduler dans ces dernières. Enfin, certaines sections particulièrement dans le développement (mais pas uniquement) sont très mouvantes en terme de tonalités. Il est plus difficile d'y définir précisément un parcours tonal, nous estimons donc les parcours tonaux qui font apparaître le plus de dominantes et de toniques. Le nom des tonalités est indiqué en anglais.

Annotation des cadences La cadence est annotée sur le point d'arrivée cadentiel. Celle-ci est préférablement placée sur un temps fort de la mesure, ainsi si certaines notes de l'accord de tonique ou de dominante attendu sont retardées, nous inscrivons tout de même le point d'arrivée cadentiel sur le temps de l'impact de l'harmonie. Nous nous efforçons d'identifier les cadences comme les définit W. Caplin, c'est-à-dire en fin de phrase. Sauf si on a une DC ou une EV, on identifie qu'une seule cadence par phrase. Si la phrase est répétée, deux cadences sont identifiées. La syntaxe pour l'étiquetage de la cadence est la suivante :

$$Deg : Type (Special)$$

Deg est le degré correspondant à la tonique de la tonalité dans laquelle est jouée la cadence par rapport à la tonalité principale. *Type* est le type de la cadence suivant la définition de la section 2.3.2. La figure 2.4 montre ainsi une cadence parfaite dans la tonalité principale de la pièce (*I* : PAC). *Special* est facultatif, il est utilisé pour les cadences importantes de la forme sonate à savoir les MC, les EEC, les MC' (MC de la réexposition) et les ESC. Ainsi, les MC de premier type dans les expositions de forme sonate sont annotées *V* : HC MC.

Annotation des formes et de la forme sonate L'annotation de la forme demande plus d'explications. Si pour certains mouvements, on peut identifier la forme et délimiter la partition sans trop d'ambiguïté (les fugues, les formes "rondo strict", thème et variations et les menuets en particulier), il est plus périlleux d'annoter voire même parfois d'identifier les mouvements de forme sonate.

Comme nous l'avons déjà mentionné précédemment (section 2.4), la forme sonate est une organisation dialectique de tonalités et de thèmes. La forme sonate ne segmente donc pas uniquement la partition en "sections", et de toute manière, les avis des théoriciens divergent sur certaines sections et certaines frontières. Si chaque mouvement de forme sonate comporte ses spécificités et difficultés d'annotations, on peut néanmoins établir quelques généralités d'annotation. Nous avons décidé de suivre les principes d'analyse de Hepokoski et Darcy dans *Elements of Sonata Theory* [65] car ils nous paraissaient cohérents et simples à modéliser informatiquement.

Avant que je ne commence ma thèse, l'équipe avait commencé une réflexion sur la forme sonate. Cette réflexion fut approfondie et précisée afin de parvenir à une méthodologie systématique d'annotation suivant les *Elements of Sonata Theory* [65] de Hepokoski et Darcy (Table 4.8).

À cette annotation de la forme, on ajoute les annotations proposées par Marius Flothuis dans [49]. Les noms de chaque section sont moins précis que ceux proposés plus haut, on essaie donc de les adapter aux notations proposées. Il y a aussi quelques divergences entre ces

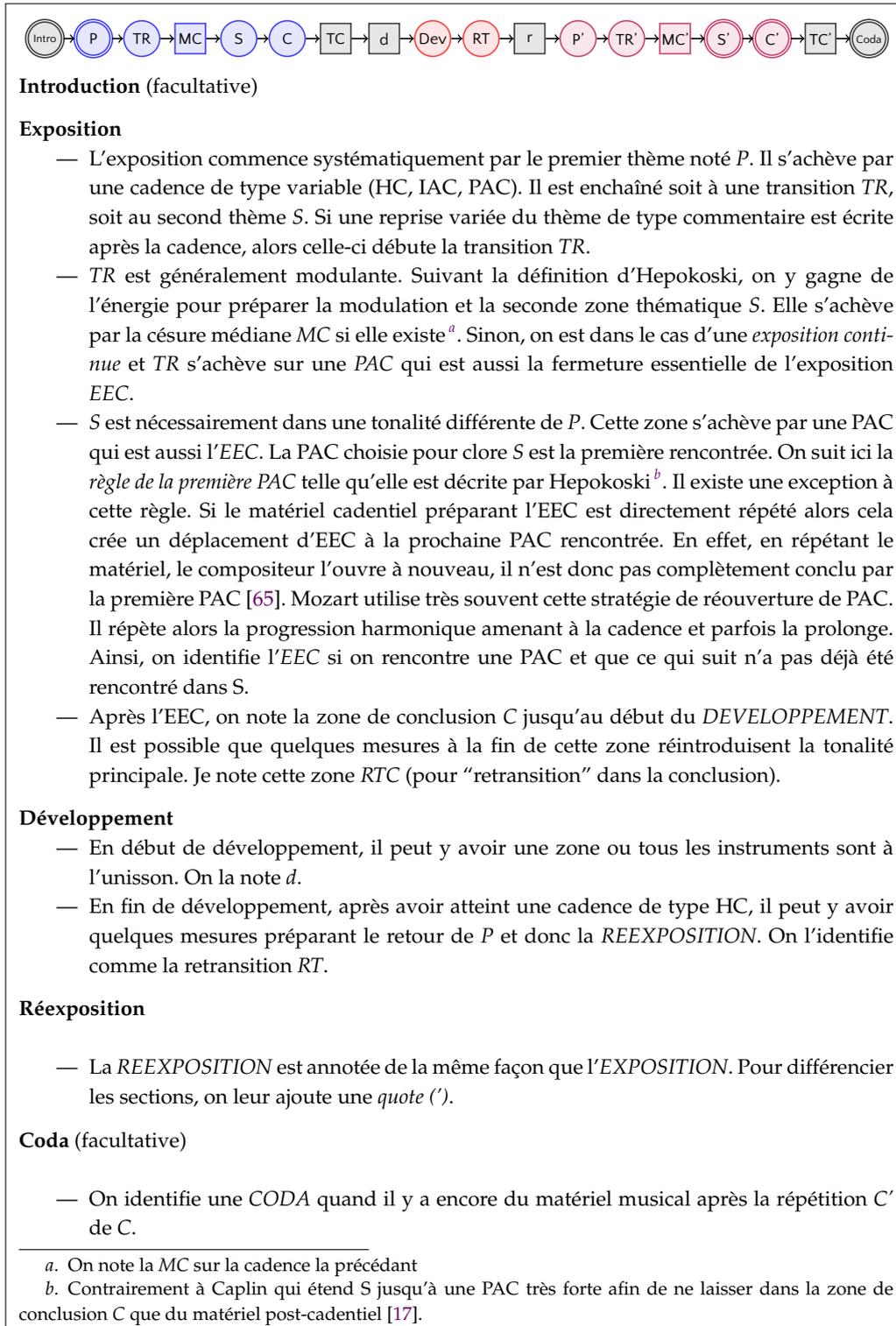


TABLE 4.8 – Méthodologie d'annotation de la forme sonate qui a été utilisée pour l'annotation des quatuors à cordes de W. A. Mozart [7].

deux annotations de référence concernant la forme de certains mouvements. Flothuis utilise plus souvent une *forme bipartite* là où nous essayons plus systématiquement de voir le schéma d'un des types de forme sonate recensé par Hepokoski. De plus, Flothuis ne semble pas utiliser la règle de la première PAC pour achever *S* ni le principe d'exposition continue.

Pour les formes menuets, la section *Menuet* et la section *Trio* sont analysées indépendamment. Pour chacune de ces sections, on observe généralement un découpage de la forme *aba'*. Pour les formes rondo, le découpage est une alternance de *refrains* et de *couplets*. Chaque couplet est identifié indépendamment (*C1C2*, etc.).

4.3.4 Utilisation du corpus

J'ai personnellement annoté l'ensemble de ce corpus, qui a été utilisé comme référence pour l'algorithme d'estimation des tonalités locales [47] que je présente au chapitre 5, ainsi que pour l'article d'équipe TISMIR estimant la forme sonate [7].

4.4 Corpus spécialisé pour séparer modulations et tonicisations : Annotations expertes et divergentes

4.4.1 Contexte

Le dernier corpus constitué est né d'une collaboration avec Néstor Nápoles López, (Université McGill, Montréal). Il est lui aussi doctorant en musicologie computationnelle et travaille également sur l'estimation automatisée de tonalité [92]. Nous avons décidé de mettre en place un corpus qui permettrait d'évaluer la performance d'un algorithme d'estimation de tonalité locale.

Ce corpus est constitué de 201 extraits musicaux annotés issus de 5 livres de théorie musicale écrits par des théoriciens de la musique (qui peuvent avoir été aussi compositeurs). Ces extraits ont été choisis car ils sont des exemples types de ce qu'est une modulation suivant les auteurs de ces livres (que nous appellerons les "théoriciens" dans la suite de cette section). Nous avons réalisé l'encodage du corpus ainsi que son annotation.

4.4.2 Références théoriques sélectionnées

Les livres choisis sont variés dans leur date d'écriture et la nationalité des auteurs. Ils sont pour certains des classiques de l'enseignement de la théorie musicale.

Aldwell, Schachter, and Cadwallader (ASC) [USA, contemporain] Les extraits de modulations sont issus du chapitre 27, *Diatonic Modulations* du livre *Harmony and Voice Leading* [6]. Ce livre a fourni 7 fichiers au corpus : quatre exemples extraits de *chorals de Bach*, un exemple du *trio pour clarinette de Mozart* et deux exemples originaux.

Kostka and Payne (KP) [USA, contemporain] Les exemples de modulations sont issus des chapitres 18 et 19 du livre *Tonal Harmony* [73]. Nous avons pris 15 exemples dans ce livre. Ces exemples sont des fragments d'œuvre de compositeurs classiques et romantiques. Il existe déjà

un CD d'exemples audio dans un livre annexe à cet ouvrage qui a notamment été utilisé dans une étude antérieure d'estimation de tonalité. Comme ils n'existaient pas sous un format symbolique, nous les avons encodés. (Figure 4.4).

G:I V⁶/₅ I IV
F: V I

FIGURE 4.4 – Exemple 18-3 dans *Tonal Harmony* de Kostka-Payne [73].

Reger (Reg) [Allemagne, 1904]

Une centaine d'exemples de modulations est présente dans le livre *On the Theory of Modulation*¹¹ [107].

Les extraits sont très courts (entre 1 et 4 mesures) et sont tous écrits par Reger lui-même. Le but du théoricien est de fournir des exemples de modulations de type cadence. Il écrit les modulations depuis *Do Majeur* vers toutes les tonalités possibles et depuis *La mineur* vers toutes les tonalités possibles. Il ajoute enfin quelques exemples pour deux tonalités extrêmement éloignées (comme depuis *Réb mineur* vers *Mi Majeur*).

Au total, Reger présente 100 exemples dans son livre mais 17 d'entre eux ont deux terminaisons : l'une est dans la tonalité majeure, l'autre dans la tonalité mineure¹². Nous avons séparé ces exemples en 2 fichiers distincts, un pour chaque terminaison. Nous avons donc 117 fichiers pour ce livre.

Rimsky-Korsakov (Rim) [Russie, 1886]

Les exemples de modulations sont issus de la troisième section du *Practical Manual of Harmony* [111]. Comme pour Reger, les 37 exemples de ce livre sont écrits par l'auteur. Cependant, certains exemples sont plus longs et plus détaillés (Figure 4.5).

Tchaikovsky (Tch) [Russie, 1872]

Les modulations sont extraites de la troisième section du *Guide to the Practical Study of Harmony* [128]. Les 25 exemples de ce livre sont écrits par Tchaikovsky.

4.4.3 Encodage et annotations

Nous avons encodé l'intégralité de ce corpus dans 201 fichiers Humdrum (**kern), un langage de représentation de musique symbolique [67]. Nous y ajoutons des annotations concernant les modulations, l'analyse harmonique et les chiffrages d'accords. Ainsi, bien que nous

11. Le livre utilisé est la republication par l'éditeur *Dover* sous le titre *Modulation* (2007).

12. Seule la tierce de l'accord terminal change.

Modulation 1
(departure key) ↓ (destination key)

C Major F Major

C:I vii^{o7}/ii ii IV/IV V/IV V⁷/IV F:I⁶ V³/₄

Modulation 2
(departure key) ↓ (destination key)

F Major C Major

I V²/V V⁶ V I⁶ V⁷/V C:I

FIGURE 4.5 – Exemple 3-17b dans le *Practical Manual of Harmony* de Rimsky-Korsakov [111]. Ici, les annotations de tonalités sont données par le théoricien. Les annotations de basse chiffrée (chiffrage arabe) ont été ajoutés dans une édition ultérieure à l'édition originale. Les annotations de degrés ont été déduites par nos soins des deux autres annotations. Les tonicisations sont encadrées.

utilisons nous-mêmes ce corpus dans un cadre d'estimation de tonalité, il peut être utilisé pour d'autres tâches. Pour ajouter les annotations d'analyses harmoniques nous utilisons la syntaxe introduite par N. Nápoles López [94].

Pour l'annotation des modulations, nous avons utilisé les annotations fournies par les théoriciens. Si ceux-ci fournissent d'autres annotations d'analyse harmonique, nous les recopions aussi, sinon nous les définissons nous-mêmes.

Concernant les tonalités, nous écrivons deux annotations pour chaque accord. L'une est associée à la modulation et correspond à l'annotation de modulation. L'autre est associée à la tonicisation et est calculée en fonction du degré harmonique et de l'annotation de modulation. À travers cette double annotation, nous voulons savoir à quel point un modèle est enclin à estimer les tonalités associées aux tonicisations ou les tonalités associées aux modulations, ce qui permet aussi d'estimer à quel point un algorithme est capable de s'extraire du contexte des notes pour "ressentir" les relations de tension/détente à grande échelle. Un exemple de cette annotation est proposé à la table 4.9.

4.4.4 Statistiques

Des statistiques sur le corpus sont présentées dans la table 4.10. Pour chaque livre, on donne le nombre de fichiers produits, le nombre de modulations, le nombre d'accords tonicisés et le nombre d'accords annotés (labels). On découpe la partition en "tranches". Une tranche est créée à chaque fois qu'une note est attaquée.

Le livre de Reger est celui qui fournit le plus de fichiers, mais aussi celui qui a le ratio le plus bas de labels par nombre de fichiers, très loin derrière les livres *ASC* (26%) et *KP* (37%). Cela vient peut-être du fait que les exemples de ces livres sont pour la plupart issus de vraies pièces musicales et qui doivent être plus longs pour bien saisir le contexte de la modulation.

TABLE 4.9 – Représentation informatique des annotations pour l'exemple de la figure 4.5. Les annotations de modulation et tonicisation sont générées d'après les annotations de tonalité et de chiffrage d'accord. Chaque ligne correspond à un accord dont la position en nombre de noires depuis le début de la partition est donnée dans la première colonne.

Position	Annotation	Modulation	Tonicisation
0	C=> :I	C major	C major
2	vii ^o 7/ii	C major	D minor
4	ii	C major	C major
6	IV/IV	C major	F major
8	V/IV	C major	F major
10	V7/IV	C major	F major
12	F=> :I6	F major	F major
14	V43	F major	F major
16	I	F major	F major
18	V2/V	F major	C major
20	V6	F major	F major
22	V	F major	F major
24	I6	F major	F major
26	V7/V	F major	C major
28	C=> :I	C major	C major

Corpus	Fichiers	Modulations	Tonicisations	Labels
ASC	7	8	7	185
KP	15	21	11	554
Reg	117	220	40	768
Rim	37	44	107	257
Tch	25	60	38	238
Total	201	555	203	2002

TABLE 4.10 – Contenu du corpus. Chacune des valeurs indique le nombre d'occurrences de chaque élément dans le corpus correspondant.

Rim et *Tch* sont les livres contenant le plus de tonicisations. On décompte des tonicisations sur respectivement 42% et 16% des accords annotés. Ces livres peuvent ainsi fortement contribuer à déterminer si un algorithme d'estimation des tonalités est plus enclin à estimer la tonalité au niveau des modulations ou des tonicisations.

Les auteurs de *Rim* et *Tch* ont aussi tendance à placer la tonalité d'arrivée lors d'une modulation sur le degré de tonique (I), alors que les autres théoriciens placent plutôt la nouvelle tonalité sur le degré de dominante (V) le précédant. Ainsi ce degré de dominante est considéré comme une dominante de dominante chez *Rim* et *Tch*, ce qui explique le nombre élevé de tonicisations.

Il est intéressant de noter cette disparité d'analyse entre les théoriciens nord-américains et allemands d'une part et les théoriciens russes d'autre part. Cette contrainte rend ce corpus inadaptable à l'entraînement d'un modèle d'apprentissage machine, c'est pourquoi nous recommandons de réserver son usage à des fins de test et d'évaluation de performance.

4.5 Bilan

La constitution, la modélisation et l'annotation de corpus sont des tâches essentielles en analyse musicale computationnelle. Des corpus de qualité sont nécessaires aux études telles que celles que nous présentons dans le reste de la thèse. Pour travailler sur la césure médiane, on a par exemple besoin d'identifier clairement quelles hypothèses musicologiques sont considérées. Ces tâches sont longues mais nécessaires et nous avons pu au cours de ce chapitre, en présenter plusieurs (Table 4.1). Une difficulté est que la *modélisation* en elle-même est un défi : c'est souvent quand on fait une première annotation qu'on voit mieux comment bien modéliser et quelles sont les catégories. Nous nous en sommes rendu compte lors du travail sur les différents travaux sur les corpus, en particulier sur celui des quatuors à cordes de Mozart qui a marqué cette thèse. Ce n'est souvent qu'après plusieurs itérations que l'on sait précisément comment on annote, ou comment on encode. Le résultat en fin de thèse est que nous disposons désormais d'une méthodologie d'annotation complète et de mouvements annotés.

Nous utilisons ces corpus pour aider à l'évaluation des modèles développés pendant la thèse à commencer par le modèle d'estimation des tonalités locales présenté dans le chapitre suivant.

Chapitre 5

Modélisation et estimation de la tonalité et des modulations

Contenu

5.1	Trois critères pour une modulation	62
5.2	Modélisation des trois critères	64
5.2.1	Ancrage dans la tonalité : évaluer la proximité d'une progression harmonique caractéristique, $c_{V \rightarrow I}(b, k)$	65
5.2.2	Compatibilité de la tonalité avec le texte musical : $d_{diat}(S, S')$	66
5.2.3	Proximité entre les tonalités : garantir une certaine stabilité tonale, $d_W(k, k')$	69
5.3	Combinaison des trois critères pour estimer le parcours tonal et les modulations	70
5.4	Corpus et méthodologie d'évaluation	71
5.5	Application (Mozart)	73
5.5.1	Discussion sur les mesures	73
5.5.2	Détection des progressions $V \rightarrow I$	73
5.5.3	Évaluation de la mesure d'ancrage de la tonalité sur K157.3	74
5.5.4	Compatibilité entre gammes	75
5.5.5	Estimation des tonalités et détection des modulations	76
5.5.6	Discussion	77
5.6	Application (modulations et tonicisations)	78
5.6.1	Modèles évalués	79
5.6.2	Discussion	80
5.7	Bilan	82

Dans la section 3.1, nous avons présenté un panorama des différentes approches utilisées en musicologie computationnelle ainsi qu'en MIR pour estimer aussi bien la tonalité principale d'une œuvre qu'une de ses tonalités locale. Nous introduisons dans cette section une approche originale ayant pour ambition de détecter précisément où les points de modulations se trouvent, à l'instar de la méthode proposée par E. Chew (voir page 36) [26].

Ces points de modulations nous intéressent car ils correspondent à des moments de basculement des attentes de l'auditeur. En effet, la tonique, note autour de laquelle sont construites les harmonies, change; les tensions se meuvent sur d'autres notes, les attentes harmoniques aussi.

En outre, les modulations sont des moments contextuels importants dans le déroulement d'une œuvre musicale que le musicologue pourra rechercher en priorité pour l'aider à identifier la structure de cette œuvre.

On appelle l'enchaînement des différentes tonalités le *parcours tonal* (section 2.2), et nous proposons, dans ce chapitre, un algorithme qui permet de l'estimer.

Pour ce faire, nous présentons quels signaux présents dans la partition permettent d'identifier une modulation. Nous en proposons une formalisation mathématique puis une implémentation. Ensuite, ces signaux sont combinés dans un algorithme dont la vocation est de détecter les modulations. Enfin, nous testons la pertinence de ce modèle à travers deux expériences, l'une sur le corpus des quatuors de Mozart (présenté section 4.3) permettant de fixer les paramètres du modèle, l'autre sur le corpus des théoriciens (présenté section 4.4) sur lequel sa performance sera comparée à celle d'autres modèles. Nous discutons pour finir les résultats obtenus par l'algorithme et proposons quelques pistes pour son amélioration.

Diffusion Le contenu de ce chapitre est la prolongation et la traduction de plusieurs travaux publiés au cours de la thèse [46, 47, 93]. C'est un travail pour lequel j'ai réalisé l'essentiel du corpus, des modèles, des expériences, de la rédaction.

5.1 Trois critères pour une modulation

Qu'est-ce qui, sur la partition, nous permet de dire qu'on a changé de tonalité? Quels aspects de la musique font ressentir à l'auditeur qu'on a modulé? Pour répondre à ces questions, nous identifions trois critères musicaux, significatifs pour la tonalité et observables sur la partition.

Un premier signal se trouve au niveau des notes elles-mêmes, porté par des *notes modulantes*. En effet, un changement de tonalité peut être identifié si *l'une des notes sur la partition ne correspond pas à la tonalité* dans laquelle nous sommes sensés être à ce moment. Cette note est caractérisée par son *altération*. Par exemple, sur la figure 5.1, le Do à la mesure 10 est une note modulante vers Fa majeur. Cependant, une note altérée ne signifie pas toujours qu'on a changé de tonalité. En effet, cette note qui n'appartient pas à la gamme de la tonalité peut aussi être ornementale, agrémenter les lignes mélodiques. On l'appelle alors une note *étrangère*. On dénombre plusieurs types de notes étrangères : *les notes de passage, les broderies, les retards, les appoggiatures, les échappées et les anticipations*.

Les relations de *tensions-résolutions* qui gouvernent la musique sont définies suivant la tonalité dans laquelle on est. Moduler, c'est donc briser ces relations, c'est changer d'état d'équi-

The image shows a musical score for the fourth movement of Mozart's String Quartet No. 15, measures 4 to 12. The score is in 6/8 time and starts in D minor. It features three V-I cadences: one in D minor (measures 4-6), one in G minor (measures 7-8, labeled 'tonicization'), and one in D minor (measures 9-10). A modulation to F major begins at measure 10. The score highlights voice leading with pink lines and specific notes with green vertical bars labeled 'a', 'b', and 'c'. A yellow highlight covers measures 10-12, which are in F major.

FIGURE 5.1 – Quatrième mouvement du quatuor à cordes No. 15 (K421) de Mozart, mesures 4 à 12.

Le mouvement est en Ré mineur. Une modulation vers Fa Majeur commence à la mesure 10. Le mouvement $V \rightarrow I$ confirme la tonalité dans les deux zones tonales. On observe également un mouvement $V \rightarrow I$ en Sol mineur à la fin de la mesure 6. C'est ce qu'on appelle un emprunt ou tonicisation. Les conduites de voix que l'heuristique cherche sont surlignées. Les ensembles de notes entendus aux moments a, b et c sont discutés dans le corps du chapitre : la présence du Do \sharp à la mesure 10 change la perception de la gamme courante entendue à ce moment-là et contribue à la perception de la modulation en Fa Majeur.

libre. Lorsqu'on est dans une tonalité donnée et dans un cas idéal, on s'attend à voir sur la partition une certaine gamme et beaucoup d'enchaînements harmoniques stabilisants, comme par exemple $V \rightarrow I$, dans cette tonalité. Sur la figure 5.1, on a, par exemple, une progression $V \rightarrow I$ en Fa majeur qui s'achève à la mesure 11. C'est ce deuxième signal que l'on va chercher à modéliser.

Ce critère est notamment mis en évidence par Nikolai Rimski-Korsakov dans son *Manuel pratique de l'harmonie* [111] :

On retient la modulation accomplie en y ajoutant une des cadences ou successions de cadence du ton suivant.

On va donc *chercher une confirmation dans l'harmonie* et chercher une stabilité sur une tonique de la nouvelle tonalité possible, voire encore mieux, un enchaînement $V \rightarrow I$. Dans la réalité, on peut avoir des différences comme, par exemple, une volonté du compositeur de retarder au maximum son enchaînement $V \rightarrow I$ pour créer un sentiment de frustration (comme expliqué sur la figure 5.2), ou bien encore, on peut broder au demi-ton inférieur pour créer de l'activité rythmique.

Enfin, en fonction des époques, un compositeur aura tendance à *privilégier la modulation vers certaines tonalités relativement à la tonalité principale*. Pour l'époque classique, les tonalités qu'un compositeur explorera seront généralement celles associées aux degrés de la gamme de la tonalité principale, avec une prépondérance pour l'exploration du ton de la dominante si la tonalité principale est majeure ou du relatif majeur lorsque la tonalité principale est mineure, comme *Fa Majeur* pour *Ré mineur* sur la figure 5.1. C'est notamment ce que l'on constate pour les quatuors de Mozart présentés à la section 4.3. Ce sera donc notre troisième critère.

Langsam und mit Ausdruck zu spielen ♩ = 88

Do Majeur

Sol Majeur ?				Do Majeur
6	5	64		2 65
I	IV	V		II V

FIGURE 5.2 – Début de la pièce 21 de l'Album für die Jugend. Le morceau débute en Do Majeur mais très vite, la compositeur amorce une modulation en Sol Majeur. En effet, les trois premiers accords de la mesure 3, chiffrés I, IV et V sont annonciateurs d'une cadence parfaite en Sol Majeur. Schumann le sait bien et décide de jouer avec nos attentes en bifurquant au dernier moment et en retournant en Do Majeur à l'aide du Fa \sharp qui affiche d'ailleurs une altération de courtoisie, preuve supplémentaire que Schumann préparait le Sol Majeur. A-t-on modulé? La question reste ouverte...

5.2 Modélisation des trois critères

Pour concevoir un algorithme capable de déterminer de façon fine quand la musique module et dans quelle tonalité la musique se dirige, il faut donc modéliser les critères présentés plus tôt de façon compréhensible pour l'ordinateur tout en gardant de la souplesse dans leur

prise en compte. Nous proposons donc d'introduire trois coûts ou distances par rapport à une tonalité k donnée pour chacun de ces critères à savoir :

- **Ancrage dans la tonalité** : a-t-on entendu dans un passé proche un enchaînement stable dans k ?
- **Compatibilité des notes** : à quel point la gamme de la partition est-elle différente de la gamme de k ?
- **Proximité entre les tonalités** : à quel point est-il difficile pour l'oreille d'appréhender le changement des relations de tensions-résolutions d'une tonalité k' vers une tonalité k ?

Nous considérons qu'il y a 42 tonalités possibles qui correspondent à tous les triplets

$$\{Do, Ré, Mi, Fa, Sol, La, Si\} \times \{\#, \natural, b\} \times \{Majeur, mineur\}$$

Pour toute partition, notre algorithme va construire un tableau D de dimension $B \times 42$, où B est le nombre total de temps de la partition. À tout temps b de la partition et à toute tonalité possible k , la valeur $D(b, k)$ estime le coût d'être dans la tonalité k au temps b en posant l'hypothèse que le parcours tonal précédent, calculé pour les temps de la partition avant b , est optimal. Considérer le temps comme unité temporelle minimale apporte une granularité fine comparable à celle induite par le rythme harmonique. On ne peut moduler que quand il y a un changement d'harmonie, cette granularité est donc cohérente.

Les sous-sections suivantes (5.2.1, 5.2.2 et 5.2.3) présentent une modélisation de ces mesures de distances.

La section 5.3 présente leur combinaison en une seule mesure $D(b, k)$. Une fois toutes les valeurs de $D(b, k)$ calculées, on explique comment chercher le chemin optimal dans le tableau par programmation dynamique. Ce chemin est le parcours tonal de la partition.

5.2.1 Ancrage dans la tonalité : évaluer la proximité d'une progression harmonique caractéristique, $c_{V \rightarrow I}(b, k)$

On cherche à modéliser la proximité d'un enchaînement $V \rightarrow I$ dans une tonalité k à un temps b . Comme décrit dans la section précédente 5.1, on souhaite que le coût d'être dans une tonalité donnée soit minimal au sens harmonique si au temps b , il y a un enchaînement de la dominante vers la tonique dans cette tonalité.

Nous définissons donc $c_{V \rightarrow I}(b, k)$ comme la distance en nombre de temps entre b et la dernière progression $V \rightarrow I$ détectée, c'est-à-dire :

$$c_{V \rightarrow I}(b, k) = \begin{cases} 0 & \text{si on est sur un V ou un I d'une} \\ & \text{progression } V \rightarrow I \text{ dans la tonalité } k \text{ à } b \\ \min[c, c_{V \rightarrow I}(b-1, k) + 1] & \text{sinon} \end{cases}$$

avec c , une constante fixée arbitrairement comme borne maximale de cette distance.

Détection de $V \rightarrow I$ Plusieurs solutions existent pour détecter cet enchaînement.

La plus évidente mais aussi probablement la plus difficile à mettre en place est de résoudre tout d'abord la tâche d'analyse harmonique et plus particulièrement d'étiquetage d'accords. Ceci est un problème de recherche complexe non résolu bien que plusieurs collègues aient obtenu des résultats intéressants mais non optimaux¹ [70, 86].

1. Par non optimaux, j'entends non applicables à notre problème par manque de fiabilité.

Une autre solution possible est de ne considérer que l'analyse fonctionnelle de chaque accord (généralement représentée par un chiffre romain) plutôt que son nom exact et ainsi ne considérer que les enchaînements dominante-tonique dans cette tonalité. Une ébauche de cette stratégie sera développée plus tard au cours de la thèse lorsque nous élaborerons un descripteur pour le calcul de la césure médiane dans la section 6.4.1.

Ici, pour détecter les mouvements $V \rightarrow I$, nous allons plutôt chercher les mouvements mélodiques *obligés* (ou *résolutions*) entre les deux harmonies. On cherche notamment à ce que :

- la *septième* de l'accord de V se résolve sur la tierce de l'accord de I.
- la *tierce* de l'accord de V (aussi appelée *sensible*) se résolve sur la fondamentale de l'accord de I.

Ces mouvements sont illustrés sur la figure 5.3.

À noter que le terme de *résolution* pour ces mouvements mélodiques n'est pas anodin. En effet, pour la période musicale qui nous intéresse, la tierce (note sensible) ainsi que la septième de l'accord de dominante, sont des notes sous tension attirées par les notes de l'accord parfait sur le degré de tonique. Pour caractériser ce phénomène, F.-J. Fétis parle de *consonance appellatives* [48].

Un autre mouvement intéressant à observer est l'enchaînement des deux fondamentales des deux accords sur une même voix. Bien que non obligé car non résolution, il est très caractéristique de l'enchaînement de la *dominante* vers la *tonique*. C'est notamment celui-ci que l'on voit à la basse lors d'une cadence parfaite, élément conclusif et stabilisant la tonique par excellence. Sur la figure 5.3, il correspond au mouvement illustré par la flèche inférieure.

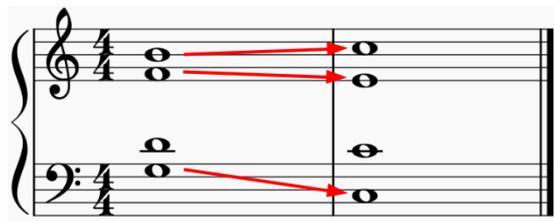


FIGURE 5.3 – Mouvements cherchés par l'heuristique pour détecter les $V \rightarrow I$. La flèche supérieure décrit un mouvement de sensible vers fondamentale. La flèche médiane décrit un mouvement de septième vers tierce. La flèche inférieure décrit un mouvement de quinte descendante (enchaînement des deux fondamentales).

Ainsi, pour chaque temps b et pour chaque voix, on regarde l'enchaînement des notes du temps $b - 1$ au temps b . Si au moins 2 des 3 mouvements décrits ci-dessus sont détectés, on considère qu'on est sur la tonique d'un mouvement $V \rightarrow I$ dans k et donc $c_{V \rightarrow I}(b, k) = 0$. On fixe également $c_{V \rightarrow I}(b_B, k) = 0$ pour tous les temps $b_B < b$ pour lesquels il y a présence de la note sensible. Sur ces temps, selon l'heuristique, nous sommes sur une harmonie de dominante.

5.2.2 Compatibilité de la tonalité avec le texte musical : $d_{diat}(\mathcal{S}, \mathcal{S}')$

On cherche dans cette section à modéliser la proximité entre ce qu'on lit dans la partition et la gamme de chaque tonalité. Avant de proposer une mesure de cette distance, j'introduis la *gamme courante* qui consiste en une représentation des notes entendues à un temps b .

La gamme courante (current diatonic pitch set) : CS

Si on compare toutes les gammes de toutes les tonalités, on remarque que la constante entre elles est le nom des notes. L'idée est donc de comparer les altérations associées à ces notes pour établir une valeur de proximité.

En effet, chaque note entendue par l'auditeur confirme ou change sa perception d'une tonalité à un instant donné. Si on est habitué à entendre un $Do\flat$, le fait d'entendre un $Do\sharp$ bouleverse la perception de la gamme actuellement jouée. Si ce $Do\sharp$ est rejoué alors, l'auditeur pensera en effet qu'on a changé de tonalité et que la nouvelle tonalité contient cette note.

Bien sûr, les approches axées sur les histogrammes présentés dans la section 3.1 prennent en compte des statistiques sur l'appartenance d'une note à une gamme d'une tonalité donnée. Cependant, ils ne considèrent pas la *directionnalité* de la musique [90]. Un $Do\sharp$ qui se résout sur un $Ré$ est un signe fort de modulation en $Ré$ Majeur ou mineur et ce peu importe le nombre de $Do\sharp$ entendus auparavant.

Pour modéliser ce phénomène, nous introduisons donc la *gamme courante* (ou *current diatonic pitch set* ou *current diatonic scale*).

On définit \mathcal{N} comme l'ensemble des sept noms de notes

$$\mathcal{N} = \{Do, Ré, Mi, Fa, Sol, La, Si\}$$

La gamme courante au temps b , notée $CS(b)$, est un vecteur de 7 valeurs associant à chacun des noms de notes de \mathcal{N} la dernière altération ($\flat, \flat, \natural, \sharp, \sharp\sharp$) entendue pour ce nom de note sur b ou avant. Si aucune altération n'a été entendue, alors l'altération considérée est celle suggérée par l'armure. La gamme courante peut donc prendre théoriquement 5^7 valeurs différentes.

Par exemple, sur la figure 5.4, au deuxième temps de la mesure 27 (temps surmonté d'un triangle blanc), la gamme courante à ce temps b_{m27b2} est :

$$CS(b_{m27b2}) = \{Do\sharp, Ré, Mi\flat, Fa\sharp, Sol, La, Si\flat\}$$

Le $Do\sharp$ apparaît dans les mesures précédant la figure.

À noter que le calcul automatique de la gamme courante ne peut s'effectuer que sur des fichiers musicaux dont l'encodage distingue les notes enharmoniques. C'est le cas par exemple des fichiers `humdrum` et `musicXML` mais pas des fichiers `MIDI`.

Gammes usuelles (ou usual diatonic pitch set)

Pour une tonalité donnée, nous nous attendons à entendre une certaine gamme. Nous définissons $\mathcal{S}(k)$ de la façon suivante :

- pour les tonalités majeures, on considère les notes de la gamme majeure usuelle. Pour Do Majeur, la gamme est donc $\{Do, Ré, Mi, Fa, Sol, La, Si\}$.
- pour les tonalités mineures, on considère les notes de la gamme mineure harmonique. Pour Do mineur, la gamme est donc $\{Do, Ré, Mi\flat, Fa, Sol, Lab, Si(\natural)\}$.

Distances entre gammes et mesure de compatibilité entre notes

Cette approche permet d'exprimer la distance entre 2 gammes de plusieurs façons :

- on exprime la disparité entre 2 gammes comme étant le nombre de notes différentes.

FIGURE 5.4 – Premier mouvement du quatuor à cordes no. 13 en Ré mineur de W. A. Mozart (K. 173) mesures 27 à 29. C’est un passage très modulant qui aboutit à une tonalité de Mi mineur, très rare et inattendue dans ce contexte et à cette époque, juste après la seconde zone thématique du quatuor à cordes. À chaque mesure, on rencontre une progression $V \rightarrow I$, d’abord en Sol mineur, puis en Do Majeur et en Si Majeur (certaines d’entre elles sont juste des emprunts dans les tonalités sus-mentionnées). Le deuxième temps de la mesure 27, discuté dans le texte, est surmonté d’un triangle blanc pour une identification aisée.

- on peut aussi rendre cette distance plus informative, en comptant une distance positive quand on gagne des dièses, mais négative quand on gagne des bémols.
- on peut même faire porter encore plus d’informations à cette distance en identifiant non seulement quel degré a changé mais aussi comment il a changé. (ex : $+4$ si le Fa en Do Majeur est devenu Fa \sharp).

Pour estimer d_{diat} , nous choisissons d’utiliser le premier modèle. Ainsi, étant données deux gammes \mathcal{S} et \mathcal{S}' , on définit la distance $d_{diat}(\mathcal{S}, \mathcal{S}')$ comme le nombre de notes différemment altérées entre elles :

$$d_{diat}(\mathcal{S}, \mathcal{S}') = |\{n \in \mathcal{N} \text{ avec } \mathcal{S}[n] \neq \mathcal{S}'[n]\}|$$

On évalue la compatibilité entre les notes de la partition au temps b et chaque gamme associée à une tonalité k en comparant les altérations entre la gamme courante $\mathcal{CS}(b)$ et les altérations de $\mathcal{S}(k)$, la gamme usuelle de k .

Pour chaque temps b et chaque tonalité candidate k , on utilise comme mesure de proximité la valeur $d_{diat}(\mathcal{CS}(b), \mathcal{S}(k))$. Par exemple, sur la figure 5.4 :

$$d_{diat}(\mathcal{CS}(b_{m27b2}), \mathcal{S}(\text{Sol mineur})) = |\{Do\}| = 1$$

$$d_{diat}(\mathcal{CS}(b_{m27b2}), \mathcal{S}(\text{Do mineur})) = |\{Do, Fa, La, Si\}| = 4$$

Cette mesure exprime que le fait d’avoir entendu un Fa \sharp fait qu’être en Sol mineur est plus pertinent qu’être en Do mineur puisque sa valeur est inférieure pour Sol mineur.

Dans la figure 5.1, il y a une modulation de Ré mineur vers Fa Majeur. Sur le point d’arrivée cadentiel en Ré mineur, au premier temps de la mesure 8, on a :

$$\mathcal{CS}(b_{m7b1}) = \{Do\sharp, Ré, Mi, Fa, Sol, La, Sib\},$$

et cette gamme courante est conservée durant les mesures 8 et 9.

On entend un Do \natural à la mesure 10. La gamme courante devient :

$$\mathcal{CS}(b_{m10b2}) = \{Do, Ré, Mi, Fa, Sol, La, Sib\},$$

ce qui est exactement la gamme de Fa Majeur.

Discussion sur la mesure choisie

Pour exprimer la mesure de compatibilité, le premier modèle a été choisi parmi les trois présentés.

En effet, bien que ce modèle contienne moins d'informations que les autres, il est tout à fait calibré pour estimer une mesure de dissemblance, et donc par extension de distance, entre deux gammes. Il est facilement manipulable et combinable aux autres distances que l'on cherche à estimer.

Bien que cela reste encore très conceptuel, le troisième modèle nous semble très intéressant pour modéliser les modulations. On pourrait par exemple imaginer ce type de formulation :

$$Sol_{Majeur} = Do_{Majeur}, +\hat{4}$$

ou encore

$$Sol_{mineur} = Fa_{Majeur}, +\hat{1}, -\hat{7}$$

pour formaliser une modulation et le rapport entre deux tonalités. C'est une piste qui pourra faire partie de travaux futurs.

5.2.3 Proximité entre les tonalités : garantir une certaine stabilité tonale, $d_W(k, k')$

On cherche à modéliser la vraisemblance de la transition d'une tonalité donnée vers une autre. Plusieurs systèmes mettant en relation les différentes tonalités ont été mis au point par des théoriciens de la musique.

Le plus connu est probablement le cycle des quintes, mentionné au chapitre 2 (Figure 2.3). On remarque que les tonalités adjacentes dans le cycle des quintes ont des gammes n'ayant qu'une note de différence si l'on suit le cycle majeur et même aucune si l'on considère l'axe des tonalités relatives (gamme mélodique descendante ou naturelle), ce qui en fait un bon indicateur de proximité des tonalités.

Lors de l'annotation du corpus des quatuors de Mozart, j'ai remarqué que le compositeur faisait relativement souvent la bascule entre le mode Majeur et le mode mineur d'une tonalité donnée (ton homonyme). Cela n'apparaît pas dans le cycle des quintes. Cependant, Jacob Gottfried Weber propose dans *Versuch einer geordneten Theorie der Tonsetzkunst* [137] une table des relations entre tonalités qui comble ce manque (voir figure 5.5). J'utilise donc cette organisation des tonalités pour estimer ma distance $d_W(k, k')$ (W pour Weber).

On définit $d_W(k, k')$, la mesure de proximité entre deux tonalités k et k' , comme la distance euclidienne entre ces deux tonalités dans la table de Weber :

$$d_W(k, k') = \min \left(\sqrt{|x_{k'} - x_k|^2 + |y_{k'} - y_k|^2}, z \right)$$

avec (x_k, y_k) et $(x_{k'}, y_{k'})$, les coordonnées des deux tonalités dans la table et $z = 10$ comme une constante "garde-fou" permettant de limiter la distance maximum entre deux tonalités.

Par exemple, $d_W(Ré\ mineur, Fa\ Majeur) = 1$ alors que $d_W(Ré\ mineur, Do) = \sqrt{2}$. Il est en effet plus fréquent à l'époque du classicisme musical de moduler depuis *Ré mineur* vers *Fa Majeur*, comme dans la figure 5.1, que vers *Do Majeur*.

TABLE OF THE RELATIONSHIPS OF KEYS.

<i>C</i>	—	<i>a</i>	—	<i>A</i>	—	<i>f</i> ♯	—	<i>F</i> ♯	—	<i>d</i> ♯	—	<i>D</i> ♯	—	<i>b</i> ♯	—	<i>B</i> ♯	—	<i>g</i> ×
<i>F</i>	—	<i>d</i>	—	<i>D</i>	—	<i>b</i>	—	<i>B</i>	—	<i>g</i> ♯	—	<i>G</i> ♯	—	<i>e</i> ♯	—	<i>E</i> ♯	—	<i>c</i> ×
<i>D</i> b	—	<i>g</i>	—	<i>G</i>	—	<i>e</i>	—	<i>E</i>	—	<i>c</i> ♯	—	<i>C</i> ♯	—	<i>a</i> ♯	—	<i>A</i> ♯	—	<i>f</i> ×
<i>E</i> b	—	<i>c</i>	—	<i>C</i>	—	<i>a</i>	—	<i>A</i>	—	<i>f</i> ♯	—	<i>F</i> ♯	—	<i>d</i> ♯	—	<i>D</i> ♯	—	<i>b</i> ♯
<i>A</i> b	—	<i>f</i>	—	<i>F</i>	—	<i>d</i>	—	<i>D</i>	—	<i>b</i>	—	<i>B</i>	—	<i>g</i> ♯	—	<i>G</i> ♯	—	<i>e</i> ♯
<i>D</i> b	—	<i>b</i> b	—	<i>B</i> b	—	<i>g</i>	—	<i>G</i>	—	<i>e</i>	—	<i>E</i>	—	<i>c</i> ♯	—	<i>C</i> ♯	—	<i>a</i> ♯
<i>G</i> b	—	<i>e</i> b	—	<i>E</i> b	—	<i>c</i>	—	<i>C</i>	—	<i>a</i>	—	<i>A</i>	—	<i>f</i> ♯	—	<i>F</i> ♯	—	<i>d</i> ♯
<i>C</i> b	—	<i>a</i> b	—	<i>A</i> b	—	<i>f</i>	—	<i>F</i>	—	<i>d</i>	—	<i>D</i>	—	<i>b</i>	—	<i>B</i>	—	<i>g</i> ♯
<i>F</i> b	—	<i>d</i> b	—	<i>D</i> b	—	<i>b</i> b	—	<i>B</i> b	—	<i>g</i>	—	<i>G</i>	—	<i>e</i>	—	<i>E</i>	—	<i>c</i> ♯
<i>D</i> b	—	<i>g</i> b	—	<i>G</i> b	—	<i>e</i> b	—	<i>E</i> b	—	<i>c</i>	—	<i>C</i>	—	<i>a</i>	—	<i>A</i>	—	<i>f</i> ♯
<i>E</i> b	—	<i>c</i> b	—	<i>C</i> b	—	<i>a</i> b	—	<i>A</i> b	—	<i>f</i>	—	<i>F</i>	—	<i>d</i>	—	<i>D</i>	—	<i>b</i>
<i>A</i> b	—	<i>f</i> b	—	<i>F</i> b	—	<i>d</i> b	—	<i>D</i> b	—	<i>b</i> b	—	<i>B</i> b	—	<i>g</i>	—	<i>G</i>	—	<i>e</i>
<i>D</i> b	—	<i>b</i> b	—	<i>B</i> b	—	<i>g</i> b	—	<i>G</i> b	—	<i>e</i> b	—	<i>E</i> b	—	<i>c</i>	—	<i>C</i>	—	<i>a</i>

FIGURE 5.5 – Table de Weber [137]. Les tonalités majeures sont notées en majuscules, les tonalités mineures en minuscules. Sur une ligne verticale, les tonalités s'enchaînent comme sur le cycle des quintes. Vers le haut, on enchaîne les sauts de quinte ascendants. Vers le bas, on enchaîne les sauts de quinte descendants. Sur une ligne horizontale, la règle est plus complexe. Si on est sur une tonalité majeure, la tonalité directement à sa droite sera la tonalité relative mineure et la tonalité à sa gauche sera la tonalité homonyme. On inverse les liens si la tonalité est mineure (relative à gauche et homonyme à droite).

À noter que cette table ne semble pertinente que pour estimer la distance entre tonalités pour la période classique. Les modulations sont de plus en plus "osées" au romantisme, privilégiant les modulations enharmoniques et chromatiques. Voir à cet effet les livres de Max Reger [107], de Tchaïkovski ou encore Rimski-Korsakov, tous trois compositeurs et théoriciens qui proposent de passer d'une tonalité à une autre en très peu d'accords. On pourra également trouver d'autres modèles de mise en relation des tonalités recensés par Marc Rigaudière [110, p. 142-156].

La distance euclidienne est utilisée, car des tonalités ici à distance $\sqrt{2}$ (Do / Mi mineur), semblent plus proches que d'autres à distance 2 (Do / La majeur). Cela n'invalide pas cependant l'utilisation d'autres distances qui restent à définir.

5.3 Combinaison des trois critères pour estimer le parcours tonal et les modulations

Les trois critères ayant été modélisés par des mesures dans la section précédente, on peut maintenant les réunir dans une même expression pour estimer le parcours tonal d'une œuvre.

Estimer un parcours tonal $k_1 \dots k_B$ d'une œuvre de B temps successifs nécessite d'identifier une tonalité locale k_b pour chaque temps $b \in [1, B]$. Le parcours tonal renvoyé par l'algorithme

minimise de façon optimale une combinaison linéaire des trois mesures sur toute la partition. On associe à chaque parcours tonal possible $k_1 \dots k_B$ un coût :

$$\begin{aligned} \Delta(k_1 \dots k_B) &= \sum_{b \in [1, B]} [\alpha \cdot c_{V \rightarrow I}(b, k_b)/c \\ &\quad + \beta \cdot d_{diat}(\mathcal{CS}(b), \mathcal{S}(k_b))/7] \\ &\quad + \sum_{b \in [2, B]} \gamma \cdot d_W(k_{b-1}, k_b)/z \end{aligned}$$

dans lequel les divisions par c , 7 , et z servent à normaliser chacune des trois mesures afin d'obtenir des valeurs entre 0 et 1 pour chacun des termes. Les poids α , β et γ servent quant à eux à pondérer l'importance relative de chacune des trois mesures.

Le parcours tonal optimal $k_1 \dots k_B$ qui minimise $\Delta(k_1 \dots k_B)$ peut être estimé par programmation dynamique car il est possible d'optimiser $k_1 \dots k_b$ comme la combinaison de l'optimisation partielle de $k_1 \dots k_{b-1}$ et d'une tonalité k_b [11].

L'algorithme construit le vecteur D de taille $B \times 42$. Pour un temps b et une tonalité candidate k , la valeur $D(b, k)$ estime la vraisemblance d'être dans cette tonalité k au temps b , en faisant l'hypothèse que le parcours tonal calculé jusqu'ici est optimal.

Ainsi, $D(1, k) = 0$ et pour tout $b \geq 2$:

$$\begin{aligned} D(b, k) &= \alpha \cdot c_{V \rightarrow I}(b, k)/c \\ &\quad + \beta \cdot d_{diat}(\mathcal{CS}(b), \mathcal{S}(k))/7 \\ &\quad + \min_{k'} [\gamma \cdot d_W(k, k')/z + D(b-1, k')] \end{aligned}$$

Quand les $B \times 42$ valeurs de $D(b, k)$ sont calculées, on choisit le parcours optimal $k_1 \dots k_B$ dans la table à rebours (*backtracking*) :

- pour le dernier temps de la partition b_B , on choisit la tonalité k_B qui minimise $D(b_B, k_B)$.
- pour le temps d'avant b_{B-1} , on choisit la tonalité k_{B-1} , qui a permis de minimiser la valeur $d_W(k_{B-1}, k') + D(b_{B-1}, k_{B-1})$.
- on réitère cette étape jusqu'au premier temps b_1 .

5.4 Corpus et méthodologie d'évaluation

Corpus et implémentation Le modèle présenté a été appliqué sur deux corpus. Ces expériences sont détaillées dans le reste du chapitre. Dans un premier temps, le corpus des quatuors à cordes de Mozart présenté dans la section 4.3 a été utilisé pour tester la validité du modèle. Il a aussi été utilisé dans le cadre de l'article [93] sur le corpus présenté dans la section 4.4.

L'algorithme a été codé à l'aide du framework python music21 [35]. La granularité temporelle choisie est la noire pour les morceaux à signature rythmique binaire et la noire pointée pour les signatures rythmiques ternaires. On considère cette granularité suffisante pour modéliser la presque intégralité du rythme harmonique des répertoires étudiés.

Comparaison exacte. Nous pouvons déjà utiliser les mesures classiques d'évaluation de classification. Pour chaque temps $i = 1 \dots N$ de la partition, on suppose qu'on a une annotation

prédite k_i et une annotation de référence ℓ_i . On peut alors, en considérant une annotation K (comme "Do majeur") calculer les valeurs suivantes :

- Vrai positifs : $TP_K = |\{i, k_i = \ell_i, k_i = K\}|$
- Faux positifs : $FP_K = |\{i, k_i \neq \ell_i, k_i = K\}|$
- Faux négatifs : $FN_K = |\{i, k_i \neq \ell_i, \ell_i = K\}|$

La précision (*accuracy*), ou score exact, est alors $Accuracy_K = TP_K/N$, et varie entre 0 (aucune annotation correctement prédite) et 1 (toutes les annotations correctement prédites). Si le calcul n'est pas sur chaque temps mais sur chaque note ou accord, on peut pondérer le calcul de la précision par la durée de chaque note ou accord d_i .

On utilise aussi la mesure F_1 : $F_{1K} = 2TP_K/(2TP_K + FP_K + FN_K)$. La mesure F_1 est la moyenne harmonique entre la précision et le rappel (qui calcule $TP_K/(TP_K + FN_K)$) et est une bonne mesure de performance d'un système.

On peut enfin calculer ces valeurs non pas sur une annotation K , mais sur l'ensemble des annotations.

Comparaison "MIREX". Cependant, quand deux tonalités prédites ne coïncident pas avec une annotation de référence, une de ces deux prédictions peut être considérée comme plus précise que l'autre et ce grâce à la relation entre les tonalités comme elle a été modélisée plus haut (section 5.2.3). Par exemple, pour une tonalité locale de *Fa Majeur*, une prédiction de *Do Majeur* bien que mauvaise reste plus acceptable qu'une prédiction en *La♯ mineur* et ce dû à la proximité des tonalités. Afin de prendre en compte cette relation et évaluer un peu plus finement la performance d'un modèle, nous introduisons un score pondéré *MIREX*² pour noter nos algorithmes.

La table 5.1 présente les poids associés aux deux processus d'évaluation utilisés dans cette expérience, rapportés à la relation entre la tonalité prédite et la tonalité présente dans l'annotation de référence en un temps donné. Ces poids sont aussi ceux utilisés pour l'évaluation des algorithmes estimant la tonalité principale dans le cadre du Music Information Retrieval Evaluation eXchange (MIREX) depuis 2005 [41].

Relation entre la tonalité prédite k et la référence ℓ	$w_{\text{exact}}(k, \ell)$	$w_{\text{MIREX}}(k, \ell)$
Même tonalité	1.0	1.0
Dominante / Sous Dominante	0.0	0.5
Relatif	0.0	0.3
Ton homonyme	0.0	0.2
Autre	0.0	0.0

TABLE 5.1 – Pondération du score suivant la relation entre les tonalités.

Le score sur l'ensemble de la partition se calcule par $score_w(k_1 \dots k_N) = \sum_{i=1}^N w(k_i, \ell_i)/N$. Si le calcul n'est pas sur N temps mais sur M notes ou accords, on pondère le score de chaque note/accord i par sa durée d_i :

$$score_w(k_1 \dots k_M) = \frac{\sum_{i=1}^M w(k_i, \ell_i) d_i}{\sum_{i=0}^M d_i} \quad (5.1)$$

2. Du nom du challenge annuel Music Information Retrieval Evaluation eXchange

Le score $score_{w_{\text{MIREX}}}$ peut donc varier entre 0 (prédiction toujours mauvaise) et 1 (prédiction parfaite). Une prédiction qui ne se tromperait que d'un pas dans le cycle des quintes (prédiction de *Do Majeur* ou *Sib Majeur* au lieu de *Fa Majeur* par exemple) aurait ainsi un score entre 0.5 et 1. Le score $score_{w_{\text{exact}}}$ revient à calculer le *Accuracy* défini ci-dessus.

Cette évaluation sera utilisée dans la section sur le corpus de modulations (section 5.6).

5.5 Application à l'estimation des tonalités et à la prédiction des modulations sur le corpus des quatuors de Mozart

Nous découpons cette première expérience en deux étapes. Tout d'abord, la détection de tonalité est effectuée sur une seule des pièces du corpus afin de discuter la pertinence des mesures proposées. Ensuite, après avoir fixé les différents coefficients associés aux critères, nous étudions les résultats obtenus sur tout le corpus.

5.5.1 Discussion sur les mesures

Nous discutons ici les résultats de l'extraction des distances sur le 3^e mouvement du quatuor K. 157. Ce mouvement est de forme rondo et ses tonalités sont identifiables sans ambiguïté. Sa tonalité principale est *Do Majeur*. Les modulations observées sont dans des tonalités voisines (la tonalité de la dominante *Sol Majeur* et la tonalité homonyme *Do mineur*). Enfin, une fois les paramètres optimaux fixés, le modèle estime la tonalité correctement pour 98.4% des temps de la partition : il ne se trompe que sur 3 temps.

Les figures 5.6 et 5.7 présentent des courbes qui correspondent à la valeur des mesures $c_{V \rightarrow I}$ et d_{diat} pour 4 tonalités données, à chaque temps b de la partition. On affiche aussi la combinaison optimale D à la figure 5.9. Pour chacune de ces figures, l'annotation de la référence sera affichée en-dessous de chaque courbe de distance (ligne épaisse pleine colorée suivant les tonalités de référence).

Les 4 tonalités choisies sont :

- *Do Majeur* (courbe bleue) : la tonalité principale du mouvement.
- *Sol Majeur* (courbe verte) : la tonalité de la dominante de *Do Majeur* (modulation dans cette tonalité de la mesure 17 à la mesure 32).
- *Do mineur* (courbe orange) : la tonalité homonyme de *Do Majeur* (modulation dans cette tonalité aux mesures 49-56 puis 61-64).
- *Do \flat mineur* (courbe rouge) : une tonalité très éloignée et improbable dans notre contexte, qui nous permettra d'observer un cas limite.

La distance d_W n'est pas illustrée par une figure car elle ne peut pas être évaluée indépendamment pour chaque tonalité. On a besoin de connaître les résultats de D pour toutes les tonalités à tous les temps afin de déterminer toutes les transitions pour cette distance. Nous n'évaluons donc pas cette distance indépendamment mais avec D dans la prochaine section.

5.5.2 Détection des progressions $V \rightarrow I$

L'annotation de référence du K157.3 comprend 38 progressions $V \rightarrow I$. L'heuristique détecte quant à elle 116 progressions $V \rightarrow I$ dont 29 vrais positifs (TP). Ces résultats sont répertoriés

key	ref	pred	TP	FN	FP	F_1
<i>Do Majeur</i>	28	34	24	4	10	0.77
<i>Do mineur</i>	4	26	2	2	24	0.13
<i>Sol Majeur</i>	4	5	3	1	2	0.67
<i>Sol mineur</i>	0	3	0	0	3	.
<i>Mib Majeur</i>	2	0	0	2	0	.
autres	0	48	0	0	48	.
(<i>Do♭ mineur</i>)	0	0	0	0	0	.

TABLE 5.2 – Nombre de progressions $V \rightarrow I$ dans l'annotation de référence (ref) du mouvement K157.3, prédites par l'heuristique (pred), ainsi que l'évaluation des vrais positifs (TP), des faux négatifs (FN) et les faux positifs (FP), et la F_1 mesure associée.

dans la table 5.2. Comme attendu, la plupart de ces 87 faux positifs (FP) sont identifiés sur des tonalités homonymes et pour des progressions $I \rightarrow IV$. Par exemple, des $V \rightarrow I$ erronés sont détectés en *Fa Majeur*, *Fa mineur*, *Lab Majeur* et en *Lab mineur* alors qu'il s'agit en fait de progressions $I \rightarrow IV$ en *Do Majeur* et en *Mib Majeur*. Si l'on ne considère les résultats que pour les deux tonalités les plus présentes dans l'œuvre, à savoir, *Do Majeur* et *Sol Majeur*, il y a finalement assez peu de faux positifs (12/39).

Ainsi, cette heuristique s'avère pertinente pour détecter des signaux relatifs aux modulations, bien qu'il y ait quand même beaucoup de "bruit".

5.5.3 Évaluation de la mesure d'ancrage de la tonalité sur K157.3

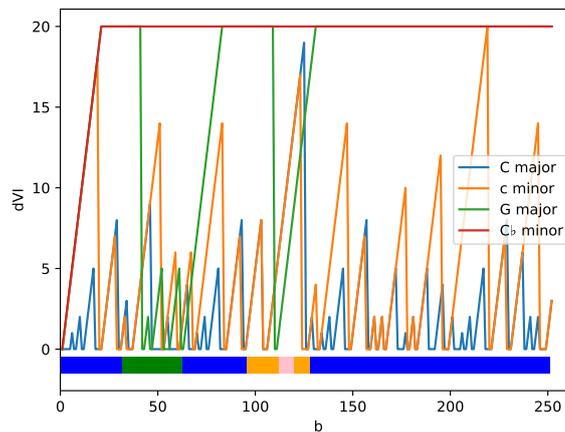


FIGURE 5.6 – Mesure d'ancrage dans la tonalité $c_{V \rightarrow I}(b, k)$ sur le troisième mouvement du quatuor K. 157 pour les tonalités $k \in \{C \text{ Majeur}, G \text{ Majeur}, C \text{ mineur}, C\flat \text{ mineur}\}$. L'épaisse barre en bas du graphe correspond au parcours tonal de l'annotation de référence.

Les estimations pour cette distance sont consultables sur la figure 5.6. La courbe *Do♭ mineur* se confond avec la courbe de *Sol Majeur* pour les premiers temps. En effet, comme aucune progression $V \rightarrow I$ n'est détectée pour ces tonalités, la courbe croît linéairement jusqu'à atteindre

la valeur $c = 20$ pour $c_{V \rightarrow I}(b, Dob\ mineur)$ pour $b > 20$, seuil fixé arbitrairement, voir la section 5.2.1. On remarque également que la courbe rouge n'est jamais réinitialisée à 0, ce qui signifie qu'on ne trouve jamais de progression $V \rightarrow I$ en *Dob mineur*. Ainsi cette mesure exclut bien les tonalités très distantes de la tonalité principale.

Notons également que les valeurs pour $c_{V \rightarrow I}(b, Do\ Majeur)$ et $c_{V \rightarrow I}(b, Do\ mineur)$ sont très souvent les mêmes. Cela permet d'identifier un premier problème : des faux positifs sont créés à cause du choix de conception de définir une progression $V \rightarrow I$ quand 2 des 3 mouvements caractéristiques sont détectés.

En effet, une remise simultanée à zéro de la distance $V \rightarrow I$ pour à la fois *Do mineur* et *Do Majeur* est possible et ce dû à la présence des mouvements de sensible vers tonique ainsi que de la fondamentale de V vers la fondamentale de I. Or la tierce de la tonique est majeure, ce qui signifie que la seule progression acceptable ici est celle en *Do Majeur*, créant ainsi un faux positif.

Il serait envisageable alors d'être plus restrictif et de n'autoriser une progression $V \rightarrow I$ que si les 3 mouvements caractéristiques sont présents, cependant cela entraînerait bien trop de faux négatifs. On touche ici à la limite des mouvements mélodiques pour détecter les progressions $V \rightarrow I$.

Cette mesure seule prédit correctement la tonalité pour 66 des 252 temps du mouvement K157.3. Elle ne peut donc pas être utilisée seule pour la détection des modulations.

5.5.4 Compatibilité entre gammes

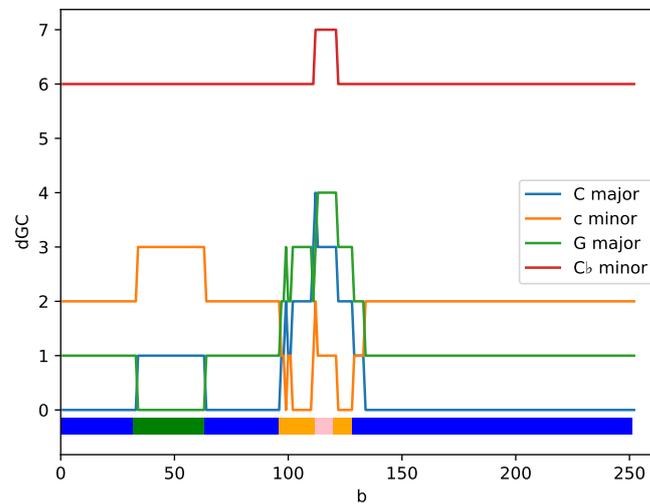


FIGURE 5.7 – Mesure de la compatibilité entre gammes ($d_{diat}(CS(b), S(k))$) sur K157.3.

Les estimations pour cette distance sont consultables sur la figure 5.7.

Comme prévu, la courbe pour $d_{diat}(b, Dob)$ est tout en haut du graphe et donc cette tonalité est bien exclue du parcours tonal par cette mesure. Pour l'immense majorité des temps (243 sur 252), cette distance est minimale pour la tonalité effectivement annotée dans la référence,

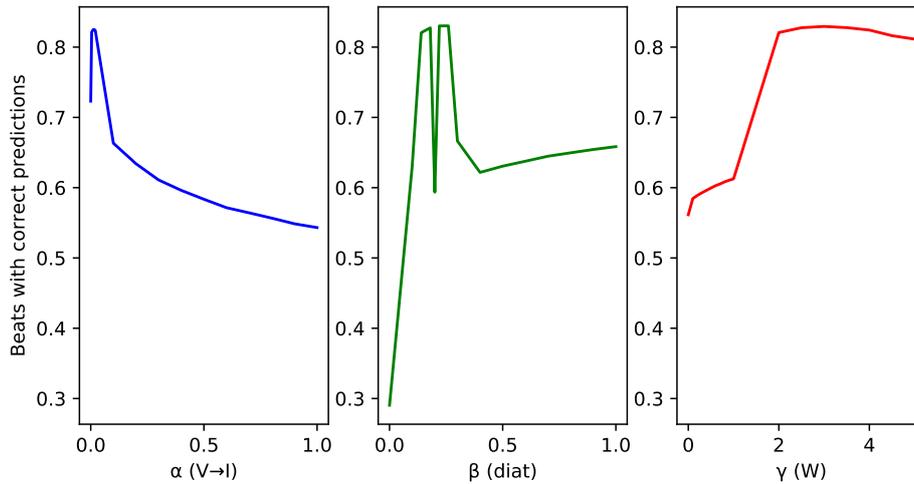


FIGURE 5.8 – Précision moyenne de l'estimation de tonalité sur les 38 mouvements du corpus. Pour chacune des courbes l'un des coefficients $\alpha/\beta/\gamma$ varie pendant que les deux autres sont constants, à leur valeur optimale.

ce qui renforce le sentiment que la gamme entendue à un temps b est un signal fort pertinent pour la modulation.

On remarque cependant qu'à partir de $b \approx 100$, aucune valeur de $d_{diat}(b, k)$ n'est à 0. À cette frontière entre deux tonalités, Mozart module progressivement en faisant entendre l'une après l'autre des notes qui seront dans la gamme de *Do mineur*. C'est pourquoi on se retrouve temporairement dans une gamme courante qui n'est pas celle d'une tonalité. Ainsi, cette distance semble pertinente pour déterminer les zones d'incertitude tonale.

Contrairement à ce que laisse penser le graphe, il y a une tonalité qui a une distance à 0 pour les temps 112 à 119. Il s'agit de la tonalité de *Mib Majeur* comme espéré au vu de la référence.

5.5.5 Estimation des tonalités et détection des modulations

On évalue maintenant l'algorithme en combinant les trois mesures.

Les coefficients α , β et γ sont déterminés en essayant différents triplets de valeurs sur le corpus des 38 quatuors à cordes de Mozart. Les valeurs testées sont comprises entre 0 et 4 (par incrément de 0.002 pour α , 0.02 pour β et 0.5 pour γ).

L'ensemble de données n'a pas été séparé en ensemble d'entraînement et ensemble de validation mais constitue un ensemble de test dans son intégralité. Il y a un risque de surapprentissage minimisé cependant par la taille plutôt modeste du corpus et des paramètres variables du modèle.

La figure 5.8 montre la précision de l'estimation de tonalité quand deux coefficients sont constants et que le dernier varie. La table 5.3 montre que bien que la mesure sur la gamme courante donne de très bons résultats toute seule (cas $\beta = 1$), l'ajout des autres mesures améliore la prédiction des tonalités de 17.5%. Les valeurs de coefficients produisant les meilleurs résultats sont $\alpha = 0.016$, $\beta = 0.3$, et $\gamma = 4.0$. Il semble que les valeurs de $c_{V \rightarrow I}$ sont généralement très hautes et ce malgré la limitation apportée par c et la normalisation de cette valeur. Elle n'aide

méthode	α	β	γ	score
seulement d_{diat}	0	1	0	67.3
seulement $c_{V \rightarrow I}$	1	0	0	16.3
pas de modulation	0	0	1	50.0
coefficients optimaux	0.016	0.3	4	84.8

TABLE 5.3 – Pourcentage des tonalités correctement détectées sur le corpus. Quand $\alpha = \beta = 0$, alors $D(b, k)$ est minimal quand on reste dans la même tonalité; L’algorithme de référence considère donc qu’il n’y a pas de modulation et que toute l’œuvre est dans la tonalité principale.

donc pas significativement à l’estimation de tonalité, probablement à cause du nombre élevé de faux positifs (voir section 5.5.2).

La table 5.4 rentre dans le détail des résultats obtenus avec les meilleurs coefficients. L’algorithme obtient une meilleure F_1 -mesure (et donc de meilleurs résultats) lorsqu’il est appliqué sur des partitions dont la tonalité principale est majeure que lorsque la tonalité principale est mineure. C’est un résultat classique de l’estimation de tonalité qui s’explique par le fait que les notes associées aux VI^e et VII^e degré d’une tonalité mineure peuvent varier suivant la gamme utilisée par le compositeur.

L’algorithme semble par ailleurs plutôt bon pour estimer la tonalité correcte des temps dans la tonalité principale. C’est un résultat heureux puisque plus de la moitié des temps du corpus sont dans la tonalité principale.

Le choix de la table dans ce contexte est particulièrement judicieux pour modéliser la relation entre les tonalités puisque les tonalités les plus usitées par Mozart sont effectivement celles côte à côte dans la table de Weber (à l’exception de la tonalité V quand on est dans une tonalité principale mineure).

Cette table pointe également les faiblesses de l’algorithme à correctement estimer les tonalités associées à la fonction prédominante (ii, iv et IV). La raison de ce comportement est encore inconnue à ce jour et nécessite plus d’investigation.

La figure 5.9 trace la mesure combinée $D(b, k)$ (pour plus de clarté, la différence entre cette mesure et la mesure pour la tonalité la plus probable est affichée) avec les coefficients α , β , et γ optimaux.

Encore une fois, on constate que *Do♭ mineur* est exclue par la mesure combinée. Le modèle arrive à trouver la tonalité principale de *Do Majeur* ainsi que toutes les modulations (*Sol Majeur*, *Do mineur*, et – non affiché sur la figure – *Mi♭ Majeur*), et ce à plus ou moins 2 temps par rapport à la modulation annotée dans l’analyse de référence.

L’estimation des tonalités est correcte pour 249 des 252 temps de la partition, ce qui améliore l’estimation calculée par d_{diat} seule. En effet, le calcul combiné de D favorise la stabilité tonale, notamment grâce à d_W , ce qui empêche l’algorithme de détecter une modulation quand des notes étrangères apparaissent sur la partition.

5.5.6 Discussion

Les résultats de l’évaluation du modèle sur le corpus des quatuors de Mozart sont satisfaisants. La combinaison des 3 mesures de distances améliore grandement les résultats d’estimation de tonalité locale par rapport à une méthode qui ne prendrait en compte que les hauteurs

Tonalité principale majeure (I)						
key	ref	pred	TP	FP	FN	F_1
I	5801	5999	5451	548	350	0.92
V	2851	2846	2525	321	326	0.89
vi	469	410	330	80	139	0.75
IV	386	382	241	141	145	0.63
i	204	217	159	58	45	0.76
v	152	129	81	48	71	0.58
ii	130	18	9	9	121	0.12
bIII	102	84	63	21	39	0.68
others	238	248	139	109	99	–
total	10333	10333	8998	1335	1335	0.87

Tonalité principale mineure (i)						
key	ref	pred	TP	FP	FN	F_1
i	1268	1038	1012	26	256	0.88
I	544	543	462	81	82	0.85
bIII	421	654	385	269	36	0.72
V	262	234	213	21	49	0.86
iv	247	133	118	15	129	0.62
v	177	230	155	75	22	0.76
others	253	340	117	223	136	–
total	3172	3172	2462	710	710	0.78

TABLE 5.4 – Précision de l’estimation de la tonalité sur 38 mouvements issus des quatuors à cordes de Mozart. Les résultats concernant les 30 mouvements dans une tonalité principale majeure sont proposés dans le tableau du dessus. Les résultats concernant les 8 mouvements dans une tonalité principale mineure sont proposés dans le tableau du dessous. Ces résultats sont obtenus en utilisant les coefficients optimaux. Les résultats sont rassemblés par tonalité. La tonalité est exprimée suivant sa relation avec la tonalité principale de chaque mouvement en terme de degré romain. Les tonalités majeures sont en majuscules, les tonalités mineures sont en minuscules.

de notes pour prédire les tonalités locales d’une œuvre. Ces résultats pourraient être améliorés en remplaçant l’heuristique pour la détection des mouvements $V \rightarrow I$ par une méthode plus fiable. Sur les quatuors de Mozart, le critère de stabilité tonale est celui qui semble le plus pertinent pour garantir une bonne estimation des tonalités locales.

5.6 Application à l’estimation des modulations et tonicisations sur le corpus spécialisé

Nous présentons maintenant une deuxième expérience d’estimation de tonalité locale que nous avons publiée dans [93]. La performance de ce modèle est évaluée sur le corpus spécialisé dans les modulations présenté à la section 4.4. Nous comparons sa performance avec deux autres modèles d’estimation de la tonalité locale, à savoir le modèle *justkeydding* de Nápoles López [92] présenté dans la section 3.1.2 et le modèle *roman* de Micchi [86] présenté dans la section 3.1.4.

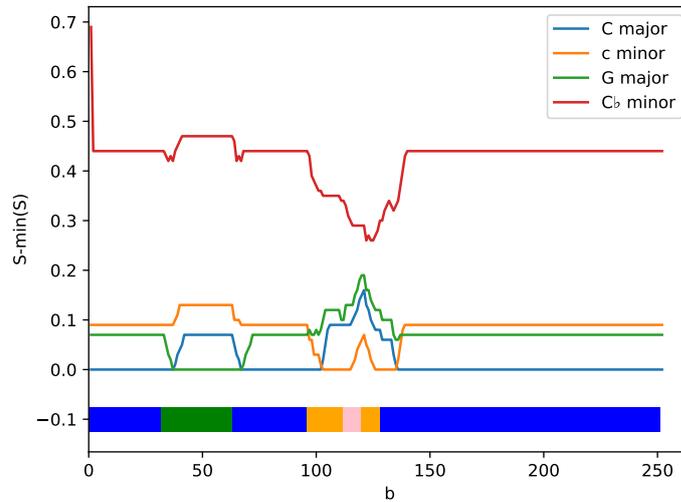


FIGURE 5.9 – Estimation de la tonalité sur K157.3 en combinant les 3 mesures avec les coefficients α , β , γ optimaux. La différence entre $D(k, b) - \min_{k'} D(k', b)$ est affichée.

5.6.1 Modèles évalués

Nous évaluons 4 modèles basiques et 3 algorithmes d'estimation de tonalité locale à l'aide de cette méthodologie.

Modèles basiques Nous concevons 4 modèles basiques afin d'établir certaines caractéristiques pour notre corpus :

- *B1* estime à chaque temps une tonalité aléatoirement. On s'attend à ce que modèle soit très mauvais et que tous les autres modèles aient une meilleure performance.
- *B2* propose à chaque temps de la partition la tonalité principale telle qu'elle est estimée par l'algorithme implémenté dans music21 [35]. On considère pour ce modèle qu'il n'y a aucune modulation.
- *B3* propose à chaque temps de la partition la tonalité qui est notée dans l'annotation de référence de la modulation. Ceci nous permet d'évaluer la surface de tonalité commune entre les annotations de modulation et de tonicisation. Nous sommes aussi curieux de savoir si des modèles d'estimation de tonalité locale peuvent surpasser ce modèle sur les tâches d'estimation de tonicisation. On s'attend à ce que ce modèle ait un score de 1 pour les tâches de modulation et un très bon score pour les tâches de tonicisation.
- *B4* propose à chaque temps de la partition la tonalité qui est notée dans l'annotation de référence de la tonicisation. Il s'agit donc du symétrique de *B3*.

Modèles d'estimation de la tonalité locale Comme indiqué précédemment, nous soumettons trois modèles d'estimation de tonalité locale à ces évaluations :

- *M1* est le modèle *justkeydding* de Nápoles Lópes [92] basé sur des HMM et des profils tonaux.
- *M2* est le modèle construit sur l'approche présentée dans ce chapitre. On teste deux paramétrages de ce modèle, *M2a* correspond au modèle non entraîné c'est-à-dire $\alpha =$

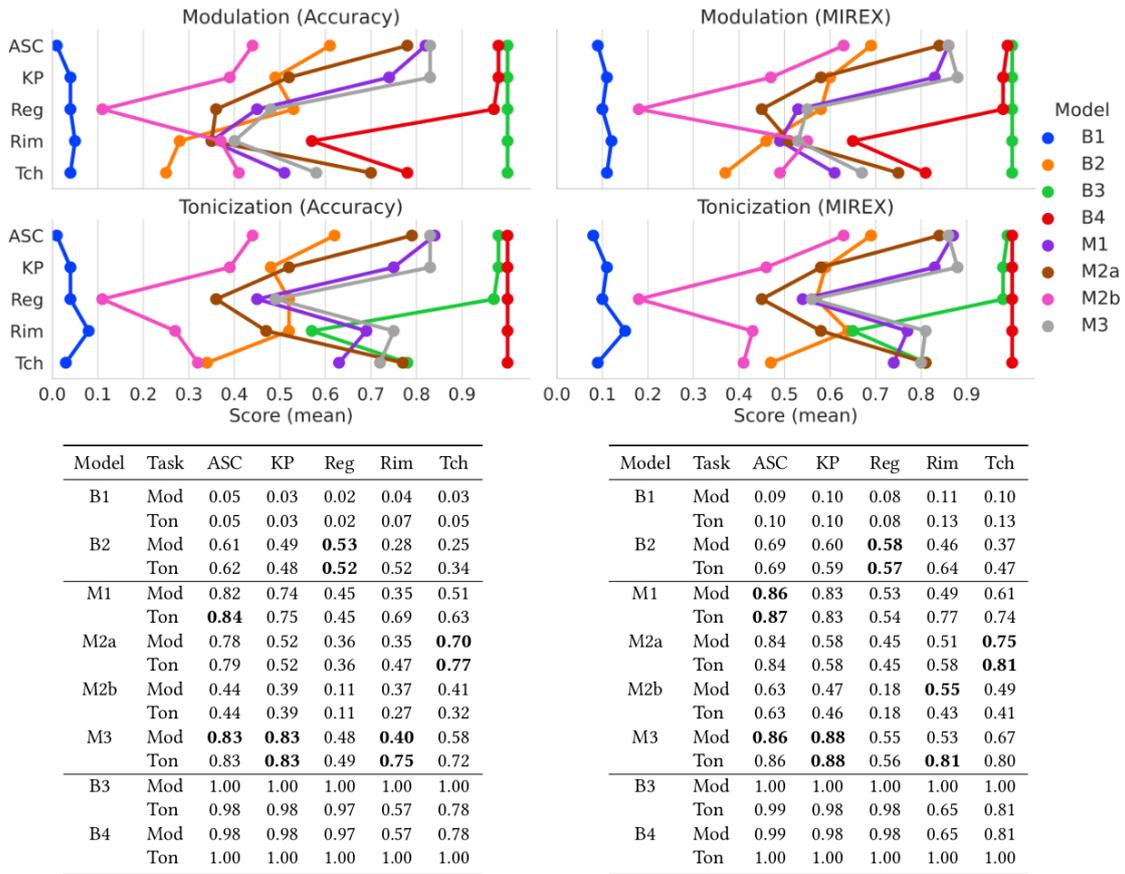


FIGURE 5.10 – Scores d'évaluation pour chaque modèle sur chaque référence théorique du corpus modulation. Les modèles prédisent les modulations et les tonicisations. De plus, ils sont évalués en utilisant la comparaison exacte (pondération w_{exact} à gauche) et la comparaison MIREX (pondération w_{MIREX} à droite), comme décrit dans la section 5.4. Une courbe est affichée pour chacune des quatre évaluations. L'axe des ordonnées montre les évaluations sur les différents livres de notre corpus, la performance variant de l'un à l'autre. L'axe des abscisses montre le score moyen obtenu par un modèle sur tous les fichiers du livre. Les scores en gras indiquent le modèle le plus performant pour un livre et une tâche donnée, en excluant B3 et B4.

$\beta = \gamma = 1$, M2b correspond aux coefficients optimaux pour l'évaluation des quatuors de Mozart présentés dans la table 5.3.

— M3 est le modèle *roman* de Micchi [86] basé sur des réseaux de neurones.

5.6.2 Discussion

Afin d'évaluer si un algorithme est plus enclin à estimer la tonalité au niveau de la modulation ou de la tonicisation, nous comparons les estimations des algorithmes aux annotations de référence correspondant aux modulations ($\ell_1 \dots \ell_2 = \text{Modulation}$) puis aux annotations de référence correspondant aux tonicisations ($\ell_1 \dots \ell_2 = \text{Tonicisations}$) obtenues comme expliqué à la section 4.4.3, suivant les deux scores $score_{w_{\text{exact}}}$ et $score_{w_{\text{MIREX}}}$. Les résultats sont présentés figure 5.10.

La figure 5.10 montre les quatre évaluations pour les modèles basiques et les modèles d'estimation de la tonalité locale. Les scores MIREX sont logiquement plus élevés puisqu'ils récompensent une prédiction fautive mais proche de la tonalité de l'annotation de référence. Les modèles qui bénéficient le plus de l'évaluation MIREX sont *B2* et *M2b*. Ces modèles privilégient la stabilité tonale (en ne modulant pas (*B2*) ou en privilégiant le fait de ne pas moduler (*M2b*)) et ce résultat montre que les modulations se font généralement dans certaines tonalités privilégiées prévisibles.

Pour les modèles *B3* et *B4*, les résultats sont symétriques comme nous l'attendions. La performance est excellente pour les exemples issus de *ASC*, *KP* et *Reg* (presque tous les labels communs) mais bien moins bonne pour *Tch* (80% de labels communs) et *Rim* (60% de labels communs). Cela s'explique par le nombre de tonicisations bien plus important chez les exemples "russes".

Les modèles d'estimation de tonalité locale obtiennent de bien meilleurs résultats lorsqu'il s'agit d'estimer la tonalité de la tonicisation que celle de la modulation, à l'exception de *M2b* (jusqu'à +0.35 points d'amélioration du score pour *M3* sur *Rim*). Ceci n'était pas spécialement prévu mais ce résultat montre que l'évaluation de ces modèles est très sensible à la moindre altération de note et privilégie donc ce critère pour déterminer la tonalité locale. Cette hypothèse est corroborée par les "meilleurs" résultats de *M2b* pour la tâche associée aux modulations et l'emphase mise sur la stabilité tonale.

Le modèle *B2* offre de façon générale des performances correctes malgré la spécialisation du corpus pour les modulations. Ce modèle parvient même à battre les modèles d'estimation de tonalité locale sur les extraits *Reg*. Ce résultat s'explique par la complexité des extraits dans *Reger*. Il arrive souvent que les modèles locaux n'arrivent pas à estimer la tonalité de départ³ alors que *B2* aura toujours bon pour les premiers accords.

Le modèle *M1* obtient généralement un score très satisfaisant (entre 0.35 et 0.87) et s'avère plus efficace pour déterminer la tonalité associée aux tonicisations, au moins pour les exemples *Tch* et *Rim*.

Le modèle *M2* s'avère bien plus décevant dans ces tâches. *M2a* obtient de moins bons résultats que *M1* et *M3* (jusqu'à 0.2 points en défaveur de *M2*), excepté pour les exemples *Tch* (entre 0.1 et 0.2 points de gain en fonction des évaluations). Bien qu'il soit meilleur en tonicisations qu'en modulations, l'écart est moins prononcé que pour *M1* et *M3*.

Les résultats pour *M2b* sont plutôt mauvais excepté pour la tâche de modulation sur le corpus *Rim* avec le score MIREX. C'est aussi le seul modèle à avoir de meilleurs résultats pour la modulation que pour la tonicisation. Ceci s'explique probablement par son entraînement sur des pièces entières dans lesquelles il y a finalement peu de modulations, d'autant plus que celles-ci demandent une préparation de longue haleine. Ce paramétrage ne semble donc pas adapté à ce corpus constitué de très courts exemples trop peu contextualisés.

Les résultats de *M3* sont légèrement meilleurs que ceux de *M1* (entre 0.4 et 0.88). Il est intéressant de voir que ces deux modèles partagent le même profil de résultats sur tous les corpus malgré leurs disparités de design et de méthode. Par exemple, les deux modèles obtiennent de mauvais résultats pour la détection des modulations chez *Rim* mais sont meilleurs pour la tonicisation.

3. En particulier les modèles *M2* qui partent de la fin pour estimer la tonalité du début ce qui est une mauvaise stratégie dans ce cas.

Enfin, bien que les exemples de modulations choisis par les théoriciens aient probablement été triés sur le volet, la diversité des annotations relevées dans la section 4.4.4 est perceptible. Notre méthode d'évaluation n'aide malheureusement pas à compenser cette ambiguïté. De futurs travaux sur cette évaluation pourront être menés dans ce sens et plus de données pourront certainement aider à résoudre ce problème.

5.7 Bilan

Dans ce chapitre, nous avons présenté une approche pour estimer le parcours tonal d'une œuvre musicale en mettant en place 3 mesures visant à estimer la différence entre les contextes musicaux idéaux portés par les tonalités et celui proposé par le compositeur sur la partition.

Si ne prendre en compte que les hauteurs de notes en un instant donné, comme le font la plupart des modèles d'estimation de tonalités, semble être une bonne stratégie, nous avons montré qu'en ajoutant plus de contexte et en particulier en prenant en compte le temps musical comme une dimension importante du problème d'estimation de tonalité au niveau local, nous pouvions améliorer la performance de ces modèles.

Si ce modèle est à son aise pour estimer les tonalités sur un corpus qui s'inscrit dans un temps musical long et complet, il ne semble pas adapté à la détection de modulation sur des extraits trop courts (selon les corpus, entre 11% et 84% de prédictions correctes).

Le fait d'appliquer un coût pour un changement de tonalité a clairement amélioré la performance de notre approche notamment lorsqu'il s'agit d'estimer des pièces de musique tonale de l'époque classique (85% de prédictions correctes sur le corpus Mozart). En effet, dans ces œuvres, le compositeur semble privilégier la stabilité tonale et moduler demande une préparation. On rejoint ici le phénomène d'accumulation d'énergie qu'Hepokoski suggère pour la modulation de la tonalité principale vers la tonalité secondaire dans l'exposition de la forme sonate [65]. Au contraire, cette contrainte plombe les résultats pour les très courts extraits dans notre deuxième expérience (selon les corpus et les méthodes d'évaluation, le score varie entre 0.11 et 0.84, ce qui est inférieur au score obtenu sur le corpus Mozart).

Bien que la prise en compte de la progression harmonique ne semble pas beaucoup aider à l'estimation de tonalité de prime abord, nous ne concluons pas à ce sujet. En effet, la méthode utilisée pour détecter les progressions V vers I n'est pas du tout fiable et déclenche beaucoup de faux positifs. Il serait intéressant d'utiliser un étiqueteur d'accords tels que ceux de Ju et al. [70] ou Micchi et al. [86] dans de futurs travaux pour estimer correctement l'apport de cette distance.

Enfin, il convient peut-être avant toute chose de faire un travail préliminaire avec les musicologues pour essayer de lever l'ambiguïté que nous avons rencontrée dans les annotations de la seconde expérience, ou bien proposer une formalisation de cette ambiguïté dans le modèle lui-même. Nous pourrions pour ce faire nous appuyer sur des travaux s'étant penchés sur ces ambiguïtés, notamment [31, 72, 125, 9].

Le prochain chapitre propose une approche bien plus simple pour une estimation de la tonalité locale afin d'aider à la détection des points d'arrivée cadentiels.

Chapitre 6

Modélisation et détection des cadences

Contenu

6.1	Un problème de classification	84
6.2	Descripteurs pour les cadences	86
6.2.1	Descripteurs du potentiel accord d'arrivée de la cadence.	86
6.2.2	Descripteurs rythmiques à proximité du point d'arrivée Z	90
6.2.3	Descripteurs directement avant le point d'arrivée cadentiel et aux alentours.	90
6.2.4	Descripteurs sur le passé plus lointain de la cadence	91
6.2.5	Statistiques sur les descripteurs	91
6.3	Détection de PAC et de HC chez Bach et Haydn	93
6.3.1	Corpus et implémentations	93
6.3.2	Processus d'apprentissage	94
6.3.3	Discussion	94
6.3.4	Bilan	96
6.4	Descripteurs pour une cadence particulière : la césure médiane	96
6.4.1	Des descripteurs pour la césure médiane	97
6.4.2	Évaluation des descripteurs sur le corpus des quatuors de Mozart	98
6.5	Détection des césures médianes chez Mozart	99
6.5.1	Processus d'apprentissage	99
6.5.2	Détection des césures médianes	99
6.6	Bilan	100

Dans ce chapitre, nous présenterons tout d'abord la façon dont nous modélisons les cadences ainsi que la stratégie mise en place pour les identifier. Cette stratégie consiste dans un premier temps à identifier des descripteurs coïncidant avec les cadences qui soient *musicaux* et qui puissent être automatiquement *extraits de la partition*. Il sera alors possible d'utiliser ces descripteurs pour entraîner un algorithme d'apprentissage pour pouvoir détecter les cadences sur de nouvelles partitions. Afin de mieux comprendre et interpréter le phénomène cadentiel, nous souhaitons exploiter et confirmer (ou infirmer) les connaissances théoriques à propos des cadences. Nous ne nous soucions pas ici de la segmentation temporelle de la partition en "tranche d'accords". C'est en effet un problème de MIR complexe qui peut être limitant pour la détection de cadence. Nous privilégions plutôt une approche originale basée sur la conduite des différentes voix.

Nous détaillons dans la section 6.1 cette approche et la façon dont nous pouvons transformer le problème de détection des cadences en un problème de classification. Nous mettons ensuite en œuvre cette approche pour résoudre et évaluer deux problèmes d'annotation.

Le premier appréhende la détection des cadences parfaites et des demi-cadences. Nous présentons comment extraire de la partition des descripteurs haut niveau issus de la théorie musicale dans la section 6.2. Ces descripteurs servent de données d'apprentissage à un modèle SVM entraîné pour détecter les cadences et en particulier, les PAC, les rIAC et les HC. Nous évaluons ce modèle sur des corpus constitués de fugues de Bach et d'expositions de quatuors à cordes de Haydn dans la section 6.3.

Le deuxième problème s'intéresse à localiser la position des césures médianes dans les expositions d'œuvre en forme sonate. Nous travaillons en particulier sur des mouvements des quatuors à cordes de Mozart issus du corpus présenté à la section 4.3. De façon parallèle, nous définissons des descripteurs spécialisés pour l'identification de la césure médiane (MC) dans la section 6.4 puis, à l'aide d'un réseau de neurones que nous entraînons grâce à ces descripteurs, nous essayons de détecter les MC de façon automatisée dans la section 6.5.

Nous concluons sur la pertinence de ces approches en les comparant avec les études antérieures (notamment celle de Giraud et al. modélisant les cadences dans les fugues [56]) et contemporaines (les modèles statistiques de Duane et al. [43, 42]) dans la section 6.6.

Diffusion Ce chapitre s'inspire très librement de deux publications élaborées pendant la thèse. Je suis coauteur principal de la première, soumise pour ISMIR2018 [14] et déjà ébauchée par l'équipe avant mon arrivée. Je suis auteur principal de la seconde, publiée à ISMIR2019 [45].

6.1 La détection de cadences vue comme un problème de classification

La cadence est un processus complet. Elle est caractérisée par une progression harmonique dont le dernier accord est appelé *point d'arrivée cadentiel*¹. Il est commode lors d'une analyse de n'indiquer sur la partition que ces points d'arrivée cadentiels en leur attachant l'annotation du type de cadence rencontré, comme on peut le voir sur les figures 6.1 ou 6.3. Comme notre but ici est l'annotation de la cadence, nous ne cherchons donc qu'à identifier où sont les points d'arrivée cadentiels.

1. Et éventuellement complété par une prolongation cadentielle.

Pour identifier où se trouve un point d'arrivée cadentiel, nous faisons une première hypothèse fondamentale : *un point d'arrivée cadentiel se trouvera toujours sur la portion entière d'un temps de la partition*. On considère alors qu'il est impossible d'avoir une arrivée cadentielle sur la deuxième croche ou la troisième double croche par exemple d'un temps. Avec cette règle, chaque temps de la partition devient donc un candidat potentiel comme point d'arrivée cadentiel.

Classification des temps de la partition La partition musicale peut être vue comme une séquence de temps musicaux, pouvant chacun être décrit par des valeurs (ou *descripteurs*) rendant compte d'aspects mélodiques, harmoniques et rythmiques. On peut par ailleurs associer à chacun de ces temps une valeur binaire qui indique si sur ce temps il y a un point d'arrivée cadentiel (valeur 1) ou non (valeur 0). Ainsi, identifier une cadence sur un temps de la partition revient à estimer, à partir de diverses informations décrivant ce temps, si on lui associe la classe "0" ou la classe "1". Le problème de détection de cadences peut ainsi être reformulé comme un *problème de classification supervisée* sur chaque temps de la partition. Le classifieur peut se constituer par *apprentissage automatique* [89] sur un ensemble de données accompagnées d'annotations spécifiant leur classe. Au terme de cette phase d'*entraînement*, le classifieur est utilisé pour estimer la classe de nouvelles données.

Descripteurs haut-niveau. Ici notre ensemble d'entrée (ou vecteur d'entrée) est constitué d'informations, appelés *descripteurs*, que l'on peut calculer directement sur la partition. Chaque temps que l'on souhaite classifier doit donc faire l'objet d'un vecteur d'entrée.

Beaucoup d'approches en MIR, que ce soit en analyse ou en génération, utilisent comme descripteurs pour caractériser la partition, des éléments relativement *bas niveau* tels que les *notes* jouées au temps considéré [33, 101]. Avec suffisamment de données, un modèle d'apprentissage profond peut théoriquement apprendre *n'importe quelle fonction* – et donc s'il y a une corrélation, même éloignée, elle peut être apprise. Cependant, la taille limitée des corpus de partitions accompagnées d'annotations de cadences auxquels nous avons accès ne permet pas un tel apprentissage. Afin de rendre possible un apprentissage pertinent, nous décidons de concevoir et d'implémenter des *descripteurs musicologiques haut-niveau* que nous pensons être pertinents pour les tâches considérées. Le principe d'utiliser des connaissances tierces (c'est à dire *non apprises*) pour transformer les données avant de les fournir au modèle, et réduire la quantité d'abstractions que doit apprendre le modèle, est connu sous le terme de *feature engineering*. Ce procédé est généralement reconnu comme une bonne pratique en apprentissage automatique car il réduit la quantité de ressources nécessaires au bon apprentissage du modèle [27].

Concevoir et évaluer des descripteurs haut-niveau permet par ailleurs de mieux comprendre sur quoi s'appuient les modèles de classification, et au final cela contribue à une meilleure explication des concepts complexes de cadence et de forme que nous étudions dans cette thèse.

Méthode d'apprentissage. Les modèles utilisés (SVM, LSTM) seront détaillés dans les sections 6.3.2 et 6.5.1. À chaque fois, pour éviter le surapprentissage, les corpus sur lesquels le modèle sera utilisé seront séparés aléatoirement en deux sous-ensembles ; un *ensemble d'apprentissage* (ou *training set*) pour entraîner le classifieur et un *ensemble de validation* (ou *validation set*) pour évaluer la performance du classifieur.

Le choix du classifieur et les valeurs de ses hyper-paramètres sont déterminés grâce à la méthode de *cross-validation Leave-One-Piece-Out (LOPO)* sur le corpus. Cette méthode consiste à évaluer la classification en entraînant le classifieur sur toutes les œuvres du corpus, excepté *une pièce* qui sert alors d'*ensemble de validation*. La validation de chaque pièce du corpus requiert l'entraînement d'un modèle dédié, constitué de l'ensemble des autres pièces du corpus.

La méthode LOPO est appréciable car elle augmente considérablement la quantité de données d'apprentissage et donc la robustesse du modèle pour les petits ensembles d'apprentissage, tel que ceux disponibles pour l'analyse musicale computationnelle. Elle est préférable à la méthode *Leave-One-Out* qui correspondrait dans notre cas à ne garder qu'*un seul temps* dans l'ensemble de validation et qui entraînerait le surapprentissage de notre modèle dû aux répétitions fréquentes que l'on rencontre dans une même pièce.

6.2 Descripteurs pour les cadences

Nous mettons en place un ensemble de 44 descripteurs binaires pour caractériser les partitions. Les descripteurs sont extraits non seulement sur le temps candidat (que nous appelons par la suite Z) mais aussi dans son passé proche. Nous espérons ainsi pouvoir repérer les progressions harmoniques caractéristiques des cadences parfaites (II-V-I, IV-V-I ou V/V-V-I) et les progressions aboutissant à la dominante pour les demi-cadences.

Nous proposons ici une heuristique simple basée sur 3 temps spécifiques : Z , $Y(Z)$ et $X(Z)$ ou plus simplement Z , Y et X . Pour un point d'arrivée cadentiel de cadence parfaite Z , Y correspond alors au premier temps de dominante V et X correspond au temps de sous-dominante II ou IV, noté SD plus haut. Une présentation de la façon dont on fixe ces points est proposée aux figures 6.1, 6.2 et 6.3 ainsi qu'à leurs légendes.

La plupart des descripteurs caractérisent des ensembles de notes jouées sur ces temps (y compris les notes attaquées avant le temps). On appellera ces ensembles $accord(Z)$, $accord(Y)$ et $accord(X)$ bien qu'ils ne soient pas forcément représentatifs de l'accord sur le temps. De plus, si certains de ces descripteurs peuvent être assimilés à de l'analyse harmonique, nous n'en faisons pas à proprement parler. En effet, il nous est inutile de savoir exactement à quel type d'accord on a affaire. Pour la détection de cadence, nous sommes surtout intéressés par la conduite de voix de la basse (enchaînement des deux fondamentales) ainsi qu'à des caractéristiques saillantes de dominantes (par exemple la présence d'une sensible) qui semblent suffisants pour caractériser une progression harmonique de type $V \rightarrow I$.

6.2.1 Descripteurs du potentiel accord d'arrivée de la cadence.

L'accord d'arrivée d'une cadence est généralement un accord parfait majeur ou mineur, suivant la tonalité à ce moment, dont potentiellement certaines des notes seront retardées ou appoggiaturées. Un premier ensemble de descripteurs identifie donc cet accord ainsi que son voisinage immédiat :

- *Z-in-perfect-major-triad* (respectivement *Z-in-perfect-triad*) : $accord(Z)$ est de la forme $\{\underline{1}, \underline{3M}, \underline{5}\}^2$.

2. Les chiffres soulignés (i.e. $\underline{1}$, $\underline{3}$, etc.) sont des hauteurs de notes calculées relativement à la basse qui n'est pas forcément la fondamentale de l'accord si celui-ci est renversé (respectivement $\{\underline{1}, \underline{3m}, \underline{5}\}$). Comme une cadence parfaite ou une demi-cadence arrive sur un accord en position fondamentale, ce descripteur est pertinent pour ces deux types de cadence

FIGURE 6.1 – Extrait du quatuor op. 17/4, i de Haydn. Une PAC est annotée à la mesure 8 (temps Z (tonique, I). Sur le temps immédiatement précédent, l'heuristique place le temps Y (dominante, V), dernier temps de la mesure 7. Le temps X est lui le deuxième de la mesure 7. Il s'agit ici du 1er renversement de l'accord de II. Les descripteurs activés détectent ici que l'accord Z est un accord parfait sous forme fondamentale et que la note la plus aiguë est la tonique, tout comme la basse (Z-in-perfect-triad, Z-in-perfect-triad-or-sus4, Z-highest-is-1). Les descripteurs indiquent aussi certaines conduites des voix entre le temps précédant Z et Z, (Z-1-comes-from-7 ①, Z-3-comes-from-4 ②), ainsi que les silences après Z (R-after-Z-rest-lowest, -middle ③) et la structure métrique (R-Z-strong-beat). Les descripteurs mettent aussi en évidence les relations avec l'accord sur Y (Y-Z-bass-moves-compatible-V-I ④, Y-Z-bass-same-voice), et sur X (X-Y-bass-moves-2nd-Maj ⑤). Notez que le choix heuristique d'un unique temps de Y implique ici que les descripteurs Y-in-V7-3 et Y-has-7 ne sont pas activés puisque l'accord Y est un accord de sixte et quarte, même si plusieurs de ces notes caractéristiques de cet accord sont effectivement présentes juste avant Z, à savoir $\underline{3}$ and $\underline{7}$ (notes entourées). Néanmoins, ces notes sont tout de même détectées par les descripteurs de conduite de voix (①, ②) ainsi que par les descripteurs sur la tonalité (Z-bass-compatible-with-I, Z-bass-compatible-with-I-scale).

- Z-in-perfect-triad-or-sus4 : $\text{accord}(Z)$ est de la forme $\{\underline{1}, \underline{3}, \underline{4}, \underline{5}\}$
- Z-is-sus4 : $\text{accord}(Z)$ est exactement $\{\underline{1}, \underline{4}, \underline{5}\}$

Ces deux descripteurs caractérisent un accord parfait majeur dont la tierce est retardée. Dans ces cas, la tierce est bien généralement préparée par la quarte qui est la septième de l'accord de V pour une cadence parfaite, mais aussi la septième de l'accord de II et V/V pour une demi-cadence.

- Z-highest-is-1 (respectivement Z-highest-is-3) : La note de la voix la plus aiguë est la tonique $\underline{1}$ (respectivement la tierce majeure ou mineure $\underline{3}$), ce à quoi on s'attend pour une PAC (rIAC).

Un autre ensemble de descripteurs décrit les mouvements mélodiques entre les notes précédant le temps Z et les notes sur le temps Z. Z- β -comes-from- α signifie que la note $\underline{\beta}$ dans l'accord(Z) est une résolution de $\underline{\alpha}$. Par exemple, l'activation du descripteur Z-3-comes-from-4 est conditionné à la présence de la hauteur $\underline{3}$ dans l'accord(Z) qui est directement précédée par la hauteur $\underline{4}$ (ce qui correspond à la résolution de la septième de l'accord de V sur la tierce de l'accord de I).

(7)

FIGURE 6.2 – Haydn, op. 17/5, i, HC au troisième temps de la mesure 8 (point Z). Le point Y est placé au temps d'avant (deuxième temps de la m8) et correspond à une harmonie de II sous position fondamentale. Le point X est placé au premier temps de la m8 et correspond au temps à une harmonie de VI sous position fondamentale. La méthodologie adoptée dans ce chapitre force l'identification d'un point X mais on ne s'attend pas à ce que cette identification soit pertinente pour une demi cadence. Les descripteurs identifient les conduites de voix (Z-6-comes-from-7, Z-4-comes-from-5, Z-1-comes-from-2 ①) prolongées par un retard de sixte au premier violon ainsi que de la quarte au second violon. (Z-6-moves-to-5, Z-4-moves-to-3 ②). D'autres descripteurs mettent en évidence la compatibilité de la tonalité avec une HC (bass-Z-compatible-with-V, notes entourées, mais à la fois Z-bass-compatible-with-I et Z-bass-compatible-with-I-scale sont Vraies et ce dû à la présence du Do# encadré) ainsi que les mouvements de basse (Y'-Y-bass-moves-chromatic, Y-Z-bass-moves-2nd-Maj ③) et la structure métrique (R-Z-strong-beat).

FIGURE 6.3 – Bach, fugue #15 en Sol majeur BWV860, PAC à la mesure 83. La conduite de voix et les notes sont soulignées comme aux figures 6.1 et 6.2. Pour s’adapter au rythme harmonique plus soutenu, chaque croche avant Z est considéré comme un Y potentiel.

Il peut aussi y avoir des notes retardées au point d’arrivée de la cadence, comme par exemple sur la figure 6.2 (identifié par ② sur la figure). De façon symétrique à l’ensemble de descripteurs précédent, nous avons mis en place les descripteurs $Z-\alpha\text{-moves-to-}\beta$. Ainsi, l’activation de $Z-4\text{-moves-to-}3$ signifie que le pitch $\underline{4}$ appartient à l’*accord*(Z) et que ce pitch est résolu sur les cadences quand la tierce de l’accord final est retardée ou appoggiaturée par la quarte.

Enfin, deux descripteurs ont été mis en place pour vérifier la cohérence entre la tonalité à proximité de Z et l’accord présumé de tonique sur Z. Ici l’estimation de la tonalité reste très sommaire contrairement à ce qui a été développé un peu plus tôt au chapitre 5 au vu de la difficulté du problème ainsi que de la précédence temporelle des travaux sur la cadence.

- *Z-bass-compatible-with-I* (resp. *Z-bass-compatible-with-V*) : est activé si les notes $\underline{4}$ et $\underline{7}$ de la tonalité qui serait effective en considérant la basse sur Z comme tonique I (resp. V) sont présentes dans les 4 temps avant Z. En effet, pour les tonalités majeures, ces deux notes n’apparaissent ensemble que dans une seule gamme pour une seule tonalité. Par exemple, en Do majeur, la note $\underline{4}$ est Fa et la note $\underline{7}$ est Si. Pour toutes les tonalités majeures, il n’y a que dans la gamme de Do majeur qu’on a à la fois Fa et Si. Ce descripteur n’est cependant pas valable pour les tonalités mineures.
- *Z-bass-compatible-with-I-scale* : est activée si on trouve toutes les notes de la tonalité sous entendue sur Z sur les 8 temps précédant Z – David Temperley suggère que la présence de l’harmonie de sous-dominante rend la PAC plus conclusive car elle fait entendre les notes de la gamme pour cette tonalité [130]. Nous espérons ainsi pouvoir identifier les tonalités mineures.

Par exemple, sur la cadence parfaite de la figure 6.1, *Z-bass-compatible-with-I* and *-with-I-scale* sont activés. Sur la demi-cadence de la figure 6.2, *Z-bass-compatible-with-V* est activé. Cependant, ces descripteurs peuvent être déclenchés par d’autres événements. Par exemple, *Z-bass-compatible-with-I* et *-with-I-scale* peuvent être activés par une harmonie de V/V^3 , ou en mineur, si Z est en fait un III en position fondamentale.

3. ce qui arrive sur la figure 6.2, dans cette figure, il y a une demi-cadence en Sol Majeur au troisième temps de la mesure 8, lors de l’estimation de la valeur du descripteur on cherche donc si toutes les notes de Ré Majeur sont présentes (puisque la basse à Z est Ré). En Sol Majeur, on n’a pas de Do# ce qui ne devrait pas activer ce descripteur. Cependant, la présence de l’harmonie V/V au troisième temps de la mesure 7 laisse entendre un Do# ce qui active le descripteur.

6.2.2 Descripteurs rythmiques à proximité du point d'arrivée Z

La cadence indique une fin de phrase musicale ainsi on s'attend à ce qu'il y ait une vraie respiration après que le point d'arrivée cadentiel ait été articulé.

Les descripteurs suivants scrutent la texture rythmique de l'œuvre afin de détecter les silences ou la continuité musicale.

- *R-Z-strong-beat* : Z est un temps fort (temps 1 et 3 pour 4/4, et 1er temps pour les autres signatures rythmiques).
- *R-Z-same-rhythm-1* (resp. *R-Z-same-rhythm-2*) : Les notes présentes un temps (resp. deux temps) avant Z sont du même nombre et ont la même durée que celles présentes un temps (resp. deux temps) après.
- *R-Z-sustained-note* : Au moins une note présente sur Z a été jouée avant Z .
- *R-after-Z-rest-highest*, *R-after-Z-rest-lowest*, *R-after-Z-rest-middle* : Il y a un silence sur la voix la plus aiguë, la voix la plus grave ou une des voix intermédiaires au temps Z . Par exemple sur la Figure 6.1, après Z au deuxième temps de la mesure 8, il y a un silence aux deux voix intermédiaires et à la voix de basse. Les descripteurs *R-after-Z-rest-middle* et *R-after-Z-rest-lowest* sont donc activés au temps Z .
- *R-after-Z-one-voice-ends* : Z est le dernier temps joué de la pièce.

6.2.3 Descripteurs directement avant le point d'arrivée cadentiel et aux alentours.

Pour les cadences parfaites, on sait que juste avant le point d'arrivée de la cadence parfaite Z , on a un accord de dominante. Pour les demi-cadences, bien qu'on ne sache pas exactement quel accord on trouvera à tous les coups avant le point d'arrivée cadentiel, on s'attend à un accord à fonction de prédominante ou plus rarement un accord de tonique. Nous décidons alors d'identifier ce temps comme étant le temps Y et l'*accord*(Y) sur ce temps.

Si c'est un accord de dominante, cela peut être un signe qu'il y a une potentielle PAC sur Z . Sinon, on espère qu'il puisse aider à l'identification des demi-cadences. Bien qu'un accord de V associé à une PAC se déploie généralement sur plusieurs temps (notamment avec la sixte et quarte de cadence), associer Y à un temps unique facilite le calcul des descripteurs.

Ainsi, on propose d'identifier Y comme étant le temps le plus proche précédant Z sur lequel une note est entendue à la basse. Ce temps peut être fixé jusqu'à une mesure avant au maximum. Si la basse n'est pas jouée juste avant Z (silence au lieu d'une note), on regarde juste avant. La durée usuelle correspondant à la préparation d'une cadence dépend du *rythme harmonique* et varie suivant les styles musicaux.

Ainsi la granularité temporelle à considérer pour trouver Y dépend du corpus. L'idéal serait de prendre la granularité sous-entendue par le tactus [10] (qu'on peut considérer de façon grossière comme la pulsation naturelle d'une œuvre musicale), mais déterminer ce niveau qui n'est pas explicitement indiqué sur une partition musicale est compliqué et constitue une tâche de MIR en soi. Cette question est d'ailleurs centrale pour formaliser la théorie de Lerdaahl et Jackendoff [79].

- *Y-Z-offsets-at-most-1* : Y est au plus une noire avant Z

Certains descripteurs concernent l'*accord*(Y) :

- *Y-has-7* (resp. *Y-has-9*) : $\text{accord}(Y)$ contient $\underline{7}$ (resp. $\underline{9}$), qui est la septième mineure de l'accord (resp. la neuvième mineure)
- *Y-in-V7* : $\text{accord}(Y)$ est un accord de septième de dominante (possiblement sans tierce ou sans quinte)
- *Y-in-V7-3* : $\text{accord}(Y)$ est un accord de septième de dominante avec obligation de la présence de la tierce Majeure

D'autres descripteurs s'intéressent à la basse :

- *Y'-Ybass-moves-8ve* : La note de basse précédant Y est à la même hauteur que celle sur Y mais à une octave différente (ce qu'on peut avoir s'il y a des accords V ou V64)
- *Y'-Ybass-moves-chromatic* : La note de basse précédant Y est à distance d'un demi-ton. (Ce descripteur est spécifique aux demi-cadences.)
- *Y-Z-bass-moves-2nd-min* (resp. *Y-Z-bass-moves-2nd-Maj*)
- *Y-Z-bass-same-voice* : La basse est à la même voix sur Y et sur Z .
- *Y-Z-bass-moves-compatible-V-I* (resp. *Y-Z-bass-moves-compatible-I-V*) : Le mouvement de basse entre Y et Z est une quarte ascendante ou une quinte descendante. (PAC) (respectivement quinte ascendante ou quarte descendante, HC I-V)

6.2.4 Descripteurs sur le passé plus lointain de la cadence

De la même façon que l'on a identifié le point Y pour l'harmonie de dominante précédant le point d'arrivée, on identifie le temps X comme le temps avant Y le plus proche ayant une basse différente (octave non compris).

Ce temps est tout aussi important que Y car il fait lui aussi partie du parcours harmonique caractéristique de la cadence parfaite SD-V-I. Il est donc essentiel de le caractériser.

Comme il n'y a aucune certitude sur la nature exacte de cette harmonie, qui peut être aussi bien une harmonie de II que de IV voire de V/V, seuls des descripteurs s'intéressant au mouvement de basse ont été mis en place. Ces descripteurs s'activent lorsque les progressions harmoniques entre parenthèses sont dans la partition :

- *X-Ybass-moves-2nd-min* (V/V65-V-I ou vii/V65-V-I)
- *X-Ybass-moves-2nd-Maj* (IV-V-I ou II6-V-I)
- *X-Ybass-moves-4th* (II-V-I ou V/V-V-I)

6.2.5 Statistiques sur les descripteurs

La table 6.1 montre, pour chaque descripteur, le nombre d'activations de chaque descripteur à chaque temps qu'il soit un point d'arrivée cadentiel ou non, ainsi qu'uniquement aux temps d'arrivée cadentiels tels qu'ils sont annotés dans la référence. On trouve à côté de ce nombre en plus petit, le nombre d'activations attendues si la distribution des activations de la feature était aléatoire et uniforme (\cdot signifie 0, et non significatif). Par exemple, il y a 70 demi-cadences sur les 7173 temps qui composent le corpus des quatuors d'Haydn. Le descripteur *Y-Z-bass-2nd-min* est activé pour 35 de ces temps de demi-cadences alors qu'il est activé pour 1323 de tous les temps du corpus. Si ce descripteur s'activait de façon aléatoire, il n'aurait été activé que 12 fois sur les temps de cadence, ce qui montre sa significativité.

Pour chaque descripteur et chaque type de cadence, la p -valeur est estimée par un test de Fisher exact calculé par la librairie Python `scipy`. Ces tests sont effectués de manière indépendante.

		bach-wtc-i			haydn-quartets		
Features		beats	PAC	rIAC	beats	PAC	HC
Rhythmic features R	<i>R-Z-strong-beat</i>	1920	60* / 25	24* / 9	3126	98* / 43	70* / 30
	<i>R-Z-same-rhythm-1</i>	394	1 / 5	. / 1	1254	2* / 17	2* / 12
	<i>R-Z-same-rhythm-2</i>	176	. / 2	. / 0	448	0 / 6	1 / 4
	<i>R-Z-sustained-note</i>	2341	14* / 31	5 / 11	2521	1* / 34	8* / 24
	<i>R-after-Z-rest-highest</i>	166	14* / 2	1 / 0	501	56* / 6	10 / 4
	<i>R-after-Z-rest-middle</i>	477	22* / 6	9* / 2	1227	72* / 16	35* / 11
	<i>R-after-Z-rest-lowest</i>	194	15* / 2	13* / 0	1130	59* / 15	34* / 11
	<i>R-after-Z-one-voice-ends</i>	180	19* / 2	2 / 0	.	. / 0	. / 0
Arrival point Z	<i>Z-in-perfect-major-triad</i>	1167	43* / 15	12 / 5	2760	94* / 38	53* / 26
	<i>Z-in-perfect-triad</i>	1819	56* / 24	19* / 9	3256	97* / 44	53* / 31
	<i>Z-in-perfect-triad-or-sus4</i>	2078	62* / 27	20* / 10	3434	97* / 47	55* / 33
	<i>Z-is-sus4</i>	680	20* / 9	1 / 3	1308	14 / 18	4 / 12
	<i>Z-highest-is-1</i>	592	55* / 7	1 / 2	1765	96* / 24	19 / 17
	<i>Z-highest-is-3</i>	1488	1* / 19	21* / 7	1596	1* / 22	28* / 15
	<i>Z-bass-compatible-with-I</i>	1724	63* / 22	23* / 8	2279	98* / 31	56* / 22
	<i>Z-bass-compatible-with-V</i>	1265	8 / 16	4 / 6	1616	3* / 22	44* / 15
	<i>Z-bass-compatible-with-I-scale</i>	1902	63* / 25	22* / 9	2104	91* / 29	46* / 20
	<i>Z-1-comes-from-7</i>	663	52* / 8	15* / 3	1016	89* / 14	30* / 9
	<i>Z-1-comes-from-1</i>	180	13* / 2	1 / 0	828	9 / 11	0* / 8
	<i>Z-1-comes-from-2</i>	523	23* / 6	7 / 2	893	65* / 12	27* / 8
	<i>Z-3-comes-from-4</i>	1078	25 / 14	16* / 5	1488	72* / 20	45* / 14
	<i>Z-4-comes-from-5</i>	197	4 / 2	. / 0	291	. / 4	9 / 2
	<i>Z-5-comes-from-5</i>	153	9* / 2	. / 0	769	2 / 10	13 / 7
	<i>Z-5-comes-from-6</i>	510	1 / 6	2 / 2	495	0 / 6	9 / 4
	<i>Z-6-comes-from-7</i>	200	. / 2	. / 1	130	. / 1	. / 1
	<i>Z-2-moves-to-1</i>	57	. / 0	. / 0	90	2 / 1	1 / 0
	<i>Z-4-moves-to-3</i>	160	2 / 2	1 / 0	340	2 / 4	11* / 3
	<i>Z-6-moves-to-5</i>	138	1 / 1	. / 0	180	. / 2	8* / 1
<i>Z-7-moves-to-1</i>	7	. / 0	. / 0	105	2 / 1	1 / 1	
Point Y	<i>Y-in-V7</i>	1267	52* / 16	17* / 6	3290	81* / 45	15* / 32
	<i>Y-in-V7-3</i>	721	44* / 9	14* / 3	2413	69* / 33	14 / 23
	<i>Y-has-7</i>	554	22* / 7	7 / 2	767	66* / 10	8 / 7
	<i>Y-has-9</i>	607	1 / 8	4 / 3	486	2 / 6	5 / 4
	<i>Y-Z-offsets-at-most-1</i>	4525	63 / 60	24 / 22	5668	90 / 78	66* / 55
	<i>Y-Z-bass-same-voice</i>	4270	63 / 56	24 / 21	5297	98* / 73	70* / 51
	<i>Y-Z-bass-moves-2nd-min</i>	1313	0* / 17	0* / 6	1328	1* / 18	35* / 12
	<i>Y-Z-bass-moves-2nd-Maj</i>	880	0* / 11	. / 4	559	0* / 7	28* / 5
	<i>Y-Z-bass-moves-compatible-I-V</i>	125	1 / 1	. / 0	448	2 / 6	6 / 4
	<i>Y-Z-bass-moves-compatible-V-I</i>	512	62* / 6	23* / 2	578	95* / 7	6 / 5
<i>Y'-Y-bass-moves-chromatic</i>	1139	6 / 15	2 / 5	2050	10* / 28	33 / 20	
<i>Y'-Y-bass-moves-8ve</i>	193	29* / 2	7* / 0	522	22* / 7	6 / 5	
Point X	<i>X-Y-bass-moves-2nd-min</i>	433	2 / 5	1 / 2	1060	10 / 14	10 / 10
	<i>X-Y-bass-moves-2nd-Maj</i>	568	25* / 7	12* / 2	803	65* / 11	5 / 7
	<i>X-Y-bass-moves-4th</i>	670	11 / 8	4 / 3	1626	4* / 22	9 / 15
Total		4739	63	24	7173	99	70

TABLE 6.1 – Décompte des activations des descripteurs pour les PAC, les rIAC et les HC.

Au vu du grand nombre de tests, seuls les descripteurs avec une p -valeur inférieure à 0.001 (**gras**, *) sont considérés comme significatif, que ce soit par leur présence ou leur absence (*italique*). Par exemple, le descripteur *Y-Z-bass-2nd-min* est significativement absent pour les PACs des deux corpus ($p < 10^{-7}$) et significativement présent pour les HCs du corpus d'Haydn ($p < 10^{-8}$).

Certaines features sont significatives pour les 2 corpus, malgré les différences de style musical et d'époque. Sans surprise, les descripteurs *R-Z-strong-beat*, *Y-Z-bass-moves-compatible-V-I*, *Z-perfect-triad-or-sus4* et *Z-highest-note-is-1* sont activés à presque chaque cadence parfaite. Une seule PAC n'active pas le descripteur *Y-Z-bass-moves-compatible-V-I*. En effet, la basse passe par une autre note de l'accord de dominante avant de se résoudre sur la tonique, ce qui explique la non activation du descripteur. Les descripteurs rythmiques et de silence sont eux aussi significatifs.

D'autres descripteurs ont une significativité différente suivant le corpus étudié. Par exemple, le descripteur *R-Z-sustained-note* est absent de presque toutes les PAC chez Haydn et pourtant, il est activé pour plus de 20% des PAC chez Bach. Cela s'explique par l'écriture contrapuntique stricte des fugues, en particulier chez Bach.

Nous espérons trouver plus de *retards* sur les temps de points d'arrivée cadentiels, pour les PAC comme pour les HC. Il s'agit en effet d'un moyen de garder encore un petit peu de tension dans le discours musical avant l'ultime résolution et repos apporté par la cadence. Il semble cependant que notre intuition ait été fautive, les résultats présentés dans la table 6.1 ne mettant en évidence aucune significativité pour ces descripteurs quel que soit le corpus.

Concernant les demi-cadences, très peu de descripteurs semblent significatifs pour les identifier. On ne trouve par exemple pas de mouvement de basse caractéristique, contrairement au descripteur *Y-Z-bass-moves-compatible-V-I* très significatif pour les cadences parfaites. En effet, la progression harmonique est moins obligée pour les demi-cadences puisque seul l'accord de V final est obligatoire. L'accord précédent peut aussi bien être entre autres un I, un II ou encore un V/V suivant la fonction de la demi-cadence et les habitudes du compositeur [82].

6.3 Détection de PAC et de HC chez Bach et Haydn

6.3.1 Corpus et implémentations

Afin de tester la pertinence du modèle présenté plus haut, nous l'avons testé sur deux corpus préexistants à cette étude (voir section 4.1 et table 4.2). On ne considère que les cadences les plus représentées et sur lesquelles on peut effectivement entraîner un modèle d'apprentissage machine. Ainsi, les rIAC ne seront détectées que pour le corpus `bach-wtc-i` tandis que les HC ne seront détectées que sur le corpus `haydn-quartets`.

La définition de cadence impliquant que le point d'arrivée d'une cadence soit la fin d'une phrase musicale, on pourrait penser que le dernier temps d'un morceau est forcément un point d'arrivée de cadence. Ce n'est pas le cas. En effet, la cadence parfaite à la fin d'une exposition est très souvent prolongée par une section conclusive (C) de longueur variable (mais relativement courte par rapport au morceau) considérée la plupart du temps comme une prolongation cadentielle et non pas comme une nouvelle cadence. Il serait ainsi très improbable que le modèle trouve une règle triviale telle qu'une cadence se trouve à la fin de l'œuvre.

		beats	ref	TP	FP	FN	F_1
haydn-quartets	PAC	3583	51	42	28	9	0.69
(21 quatuors)	HC	3583	32	18	73	14	0.29
bach-wtc-i	PAC	2357	36	26	3	10	0.80
(12 fugues)	PAC+rIAC	2357	46	30	12	16	0.68

TABLE 6.2 – Détection des cadences sur les test sets des deux corpus avec l'ensemble des descripteurs : Nombre de temps annotés dans l'annotation de référence (ref), vrais positifs (TP), faux positifs (FP), faux négatifs (FN), et F_1 mesure.

Les fichiers musicaux ont été téléchargés au format `.krn` depuis l'URL kern.ccarh.org [67]. Pour les fugues, la musique est séparée en voix plutôt que par portées comme dans l'œuvre originale pour clavier, ce qui facilite la définition et l'interprétation des descripteurs. Il est cependant tout à fait possible d'utiliser le modèle sur les partitions originales avec 2 portées et voix mélangées dans les portées. Seules les features `after-Z-rest-*` et `Y-Z-bass-same-voice` risquent d'avoir un comportement non conforme à celui attendu.

À chaque temps Z (et ses temps relatifs Y et X), les valeurs des descripteurs introduits à la section 6.2 sont extraites par un algorithme codé en Python et à l'aide notamment de la librairie `music21` [35]. La granularité pour le temps lors de la recherche des points Y et X est la noire pour le corpus Haydn et la croche pour le corpus Bach. Les classifieurs ont été implémentés à l'aide de la librairie Python `scikit-learn` [44].

6.3.2 Processus d'apprentissage

Un classifieur linéaire de type machine à vecteurs de supports (SVM) est entraîné sur chaque ensemble d'apprentissage comme explicité à la section 6.1 afin d'estimer un hyperplan séparant ainsi l'espace des descripteurs en deux classes bien distinctes.

Comme nos ensembles de données sont déséquilibrés (environ 98% des temps ne correspondent pas à des cadences), les données ont été pondérées en faveur de la classe la moins représentée, c'est-à-dire les temps correspondant aux points d'arrivée cadentiels. Nous avons également testé d'autres algorithmes de classification tels que les *k-plus-proches-voisins* ou encore les *arbres de décisions*. Les résultats obtenus sont soit moins bons, soit similaires aux SVM, nous ne les considérons donc pas dans cette étude.

Enfin, nous n'avons pas testé les réseaux de neurones profonds car les descripteurs d'entrée sont déjà de très haut niveau et les corpus utilisés semblent trop petits pour obtenir des résultats intéressants.

6.3.3 Discussion

La table 6.2 présente une comparaison entre les prédictions de chaque classifieur sur l'ensemble de test de chaque corpus et les annotations de référence. TP signifie *true positive*, il s'agit du nombre de temps qui appartiennent à la classe "point d'arrivée cadentiel" à la fois dans la référence et dans la prédiction du modèle. TN signifie *true negative* et indique le nombre de temps qui appartiennent à la classe "pas une cadence" dans la référence et la prédiction du modèle. Les FP (*false positive*) correspondent aux temps qui ne sont pas des cadences dans la

FIGURE 6.4 – Haydn, op. 55/3, i, PACs potentielles à m67 et m71. La PAC à m67 est difficile à détecter à cause du silence à la basse. Dans leur analyse globale de la forme, Sears et. al voient la fin du second thème (appelée aussi l’EEC, pour Essential Expositional Closure par [65]) à la mesure 67 et défasse ainsi toute possibilité d’une PAC plus tard dans la section conclusive qui suit [123]. Ainsi, la PAC candidate à la mesure 71 trouvée par le modèle, est comptée comme un Faux Positif. Un débat existe pour savoir si l’EEC est à la mesure 67 (première tonique cadentielle, affaiblit cependant par le silence à la basse) ou plutôt à la mesure 71 (ressenti cadentiel augmenté par les silences après ce temps et la note de basse sur le premier temps, à la mesure 67 on considérerait alors plutôt une cadence évitée).

	haydn-quartets		bach-wtc-i	
	PAC	HC	PAC	PAC+rIAC
All features XYZR	0.69	0.29	0.80	0.68
Features YZR	0.69	0.27	0.71	0.68
Features ZR	0.59	0.24	0.52	0.34
Features XYZ	0.72	0.25	0.74	0.54

TABLE 6.3 – F_1 mesure obtenue lors de la détection des cadences sur le test set des deux corpus en utilisant différents ensemble de descripteurs.

référence mais qui sont prédits comme des points d’arrivée cadentiels par le modèle, tandis que les FN (*false negative*) sont les temps identifiés comme des cadences dans la référence mais qui ne sont pas trouvés par le modèle.

Quand tous les TP, TN, FP et FN ont été dénombrés, on calcule la F_1 mesure qui correspond à la moyenne harmonique de la précision et du rappel. Il s’agit d’une bonne mesure de la performance de notre système. Elle est comprise entre 0 et 1 et plus elle est proche de 1, plus le modèle de classification est efficace.

La détection des cadences parfaites est globalement satisfaisante, avec plus de 75% de PAC détectées sur l’ensemble du corpus avec notamment très peu de faux positifs. La précédente étude de l’équipe avait réussi à détecter 82% des cadences parfaites dans les fugues de Bach [55] mais avec plus de faux positifs vu l’approche basée sur des règles de décision codées manuellement.

Après examen des 28 temps considérés comme faux positifs (FP), il s’avère que 5 d’entre eux pourraient être en effet être considérés comme des points d’arrivées cadentiels s’ils avaient

été annotés par d'autres experts, comme par exemple, dans le premier mouvement de l'opus 55/3 d'Haydn, à la mesure 71 présentée sur la figure 6.4.

La même figure propose également un exemple de faux négatif (FN) où la présence d'un silence à la basse empêche l'activation de beaucoup de descripteurs significatifs. D'autres FP sont provoqués par des accords de tonique précédant directement des points d'arrivée de demi-cadence. Ils activent ce faisant beaucoup de descripteurs significatifs pour les cadences parfaites.

Pour le corpus Bach, la détection conjointe des PAC et des rIAC comme étant une seule classe de cadence baisse les résultats. Nous attribuons ce phénomène au trop faible nombre de cadences de ce type dans la référence, le modèle n'est donc pas entraîné efficacement pour détecter ces cadences. La détection des demi-cadences (HC) a été faite uniquement sur le corpus Haydn et n'est pas très satisfaisante. En effet, comme expliqué à la section 6.2.5, aucun descripteur n'est caractéristique de la demi-cadence. Seule la moitié d'entre elles sont détectées et beaucoup de faux positifs sont identifiés par le modèle.

La table 6.3 présente l'évolution des résultats du modèle en fonction des ensembles de descripteurs utilisés pour entraîner le même modèle. Certains descripteurs sur Z considèrent déjà le passé. Néanmoins, les descripteurs aux alentours de Y sont essentiels pour améliorer l'efficacité du modèle. Les descripteurs sur X apportent un gain de performance plus mesuré mais significatif pour les PAC. Les descripteurs rythmiques (R) améliorent la détection des HC surtout chez Haydn et la détection globale chez Bach, notamment grâce au descripteur *R-Z-strong-beat* qui filtre correctement plus de la moitié des temps de la partition.

6.3.4 Bilan

Cette approche combinant apprentissage machine et descripteurs haut niveau s'est avérée efficace pour trouver les PAC sur les partitions musicales. Nous essayons maintenant de détecter les césures médianes qui en plus d'être caractérisées par des cadences sont des marqueurs importants de la forme sonate.

6.4 Descripteurs pour une cadence particulière : la césure médiane

Pour poursuivre ce travail sur les cadences et le rapprocher des travaux sur la prédiction automatisée de la structure de la forme sonate, nous avons adapté la méthode décrite ci-dessus à la détection de la cadence précédant la césure médiane (MC).⁴

La césure médiane est un des buts cadentiels de la forme sonate selon Hepokoski [64], comme nous l'avons présenté à la section 2.4. Elle correspond à un moment de rupture de la musique entre la transition TR et la seconde zone thématique S . Puisqu'elle est un moment important de l'exposition, la césure médiane est préparée par plusieurs signaux musicaux.

Nous étudions dans cette même section les descripteurs implémentés pour détecter les signaux permettant d'identifier la césure médiane présentés à la section 2.4.3.

4. Par abus de langage, nous désignons aussi la cadence précédant la césure médiane comme la césure médiane.

Nous essayons ensuite de détecter les césures médianes dans 27 mouvements de forme sonate issus des quatuors de Mozart (corpus décrit à 4.3) à l'aide d'un réseau de neurones de type LSTM. Ceci constitue une expérience dont nous discutons les résultats dans la sous-section 6.5.

6.4.1 Des descripteurs pour la césure médiane

À l'aide des descriptions de ces signaux, nous extrayons à chaque temps un vecteur de 13 valeurs correspondant à 13 descripteurs qui nous semblent pertinents pour estimer la présence des césures médianes.

La plupart des signaux présentés ci-dessus sont bien souvent trop complexes à formaliser tels quels, ils sont même parfois des sujets de recherche à part entière qui ont à peine été esquissés⁵. Ainsi nous mettons en place des descripteurs qui permettent d'observer des comportements qui peuvent se rapporter aux signaux observés mais qui ne décrivent pas exactement ces signaux (sauf si ceux-ci sont observables sans ambiguïté).

Ces descripteurs formalisent la texture tant au niveau rythmique que mélodique, mais aussi l'harmonie et le contexte tonal de la partition. Pour chaque temps b , on estime les valeurs normalisées entre 0 et 1 des descripteurs suivants :

f-rhythm-density compte la densité rythmique, c'est-à-dire le nombre de notes dans l'intervalle de temps $[b, b + 1[$ (changement de texture rythmique).

f-rest compte le nombre de silence au temps b (interruption du texte musical, la césure à proprement parler)

f-mean-pitch est la valeur MIDI moyenne des hauteurs de notes dans l'intervalle $[b, b + 1[$ (changement de texture mélodique).

f-range-pitch est la différence entre la valeur MIDI maximale et la valeur MIDI minimale des hauteurs de notes comprises dans $[b, b + 1[$ (changement de texture mélodique).

f-time est la valeur du ratio entre le temps courant et le nombre total de temps de la partition (position temporelle).

f-hammer-blow regarde si on est sur un deuxième ou un troisième coup de marteau au temps b (caractéristique de la MC comme sur la fig 2.9).

f-maj-tonic, *f-min-tonic*, *f-predominant*, *f-dominant-of-dominant* et *f-dominant* sont des descripteurs dont la valeur indique la pertinence d'associer une fonction harmonique à l'ensemble des hauteurs de notes présentes sur b . On définit pour chaque fonction des notes pertinentes qui augmentent la valeur du descripteur et des notes non pertinentes qui la diminuent. En voici la liste :

	pertinent $h(p) = 1$	non pertinent $h(p) = -1$
<i>f-maj-tonic</i>	1, 3, 5	♯4, 7
<i>f-min-tonic</i>	1, ♭3, 5	♯4, 7
<i>f-predominant</i>	2, 4, ♯4, ♭6, 6, 1, ♭2	3, 5, 7
<i>f-dominant-of-dominant</i>	2, ♯4, ♭6, 6, 1, ♭3	3, 4, 5, ♯5, 7
<i>f-dominant</i>	5, 7, 2, 4, 6, ♭6	1, ♯5

5. La modélisation de la texture est un sujet que l'équipe Algomus commence tout juste à appréhender par exemple.

	P	TR	MC	S	C
<i>f-rhythm-density</i>	0.442	0.517	0.524	0.557	0.577
<i>f-hammer-blow</i>	0.080	0.087	0.143	0.070	0.059
<i>f-rest</i>	0.223	0.194	0.099	0.229	0.181
<i>f-mean-pitch</i>	0.770	0.768	0.786	0.782	0.766
<i>f-range-pitch</i>	0.373	0.434	0.508	0.420	0.441
<i>f-time</i>	0.053	0.150	0.189	0.270	0.350
<i>f-maj-tonic</i>	0.679	0.621	0.593	0.575	0.595
<i>f-min-tonic</i>	0.650	0.598	0.581	0.561	0.588
<i>f-predominant</i>	0.568	0.553	0.576	0.545	0.521
<i>f-dominant-of-dominant</i>	0.493	0.537	0.594	0.561	0.525
<i>f-dominant</i>	0.629	0.670	0.737	0.699	0.719
<i>f-current-scale-diff</i>	0.065	0.182	0.262	0.228	0.237
<i>f-current-scale-relative</i>	0.008	0.069	0.094	0.146	0.134

TABLE 6.4 – Valeur moyenne des descripteurs selon la section de l'exposition.



FIGURE 6.5 – Distribution de deux descripteurs pertinents pour l'identification des césures médianes. Du bas vers le haut, P (marron), TR (bleu), MC (rouge), S (vert), and C (violet).

f-current-scale-diff compte le nombre de valeurs différentes entre la *gamme courante* (voir section 5.2.2) au premier temps et celle au temps b (changement de tonalité).

f-current-scale-relative compte aussi le nombre de valeurs différentes entre la *gamme courante* au premier temps et celle au temps b mais pondère ces changements. Si on a un \sharp en plus alors la valeur est +1 sinon si c'est un \flat la valeur est de -1 (changement de tonalité, en particulier de la tonique vers la dominante).

6.4.2 Évaluation des descripteurs sur le corpus des quatuors de Mozart

Les vecteurs sont extraits sur 27 expositions de mouvements identifiés comme étant de forme sonate dans le corpus de quatuors de Mozart 4.3. Les valeurs moyennes en fonction de la section de l'exposition et du descripteur considérés sont présentés à la table 6.4. La figure 6.5 détaille, quant à elle, la distribution de deux descripteurs particuliers.

Certains descripteurs ont de plus grandes valeurs sur la MC, notamment *f-dominant-of-dominant* and *f-dominant*. Comme attendu, les descripteurs visant à caractériser une fonction harmonique de tonique ont de plus grandes valeurs dans les sections P et TR tandis que les descripteurs sur les harmonies à fonction de dominante sont plus grands sur la MC ainsi que dans les sections S et C.

Les descripteurs sur la *gamme courante* ont des valeurs plutôt faibles sur P ce qui témoigne d'une stabilité tonale initiale. Ceux-ci s'activent peu à peu tout au long de l'exposition au fur et à mesure qu'on module vers la tonalité secondaire. *f-current-scale-diff* est maximal sur la MC et juste avant. Ce résultat reflète les oscillations harmoniques et l'instabilité tonale rencontrées dans ces sections. S étant le lieu de la nouvelle tonalité, la valeur de *f-current-scale-diff* diminue un petit peu. Pour *f-current-scale-relative*, la courbe est un peu différente. Le fait que cette valeur

soit constamment croissante montre qu'on peut aussi bien gagner des \flat que des \sharp dans TR et sur la MC. Ceci peut témoigner d'une instabilité modale relativement fréquente comme on peut le constater sur la figure 2.9. Les autres descripteurs pertinents sur la MC sont, comme attendus, *f-hammer-blow* ainsi que *f-range-pitch*. Ce dernier descripteur est maximal grâce à l'ouverture obtenue par le saut d'octave descendant souvent rencontré à la répétition d'un accord plaqué de type coup de marteau.

6.5 Détection des césures médianes chez Mozart

6.5.1 Processus d'apprentissage

Pour identifier automatiquement la césure médiane dans notre corpus de 27 expositions issues des quatuors de Mozart, nous entraînons un réseau de neurones récurrent de type Long Short-Term Memory (LSTM) sur ce corpus en utilisant la stratégie *Leave-One-Piece-Out*, toujours pour des raisons de taille du corpus. Le type LSTM a été choisi car il est muni de connections de rétroactions ce qui le rend non seulement particulièrement adapté sur des données organisées en séquences temporelles telle que la vidéo ou la musique, mais également intéressant pour notre problème de classification car nos descripteurs modélisent ce qu'il se passe sur le temps de MC mais aussi sur la préparation de cette MC.

Pour chaque temps b , le LSTM reçoit les valeurs des 13 descripteurs à ce temps b mais aussi aux 4 temps précédents et aux 4 temps suivants, ce qui permet de fournir au modèle une vision relativement large de ce qu'il se passe à proximité de ce temps. Lors de la phase d'entraînement, on associe également à ce vecteur d'entrée une étiquette qui vaut 1 si une MC est présente sur le temps considéré ou les 5 temps suivants, 0 sinon. Le réseau de neurones a été implémenté à l'aide du framework Python Keras [28].

6.5.2 Détection des césures médianes

À chaque temps de la partition dont on essaie de détecter la césure médiane, le modèle renvoie une probabilité que ce temps soit effectivement le moment de la césure médiane. Ainsi, la MC est positionnée au temps pour lequel le modèle renvoie la plus grande probabilité relativement au reste du mouvement. Nous considérons qu'une prédiction est correcte si celle-ci est dans un intervalle de plus ou moins 4 temps autour de la MC dans l'annotation de référence, ce qui nous semble raisonnable étant donnée la difficulté de modélisation de ce phénomène.

La figure 6.6 affiche la position de la MC dans l'annotation de référence pour chaque pièce du corpus. Le réseau en identifie correctement la position de 14 des 27 mouvements, ce qui est une amélioration par rapport à la précédente étude cherchant les MC [7] dans laquelle le modèle mis en place n'en trouve que 8. Étant donnée la taille relativement réduite du corpus, nous n'identifions pas de corrélation entre la précision de la prédiction et le mode de la pièce, le tempo ou encore le type de MC identifié. Par exemple, la position des césures médianes est correctement estimée pour 11 des 21 pièces majeures et 3 des 6 pièces mineures. Même chose pour le tempo avec 9 MC trouvées pour 18 pièces à tempo rapide et 5 MC trouvées pour 9 pièces à tempo lent.

La figure 6.7 détaille l'estimation de la probabilité d'avoir une MC sur un temps pour 4 œuvres du corpus. Pour deux de ces pièces, le modèle est d'une efficacité redoutable. Sur le

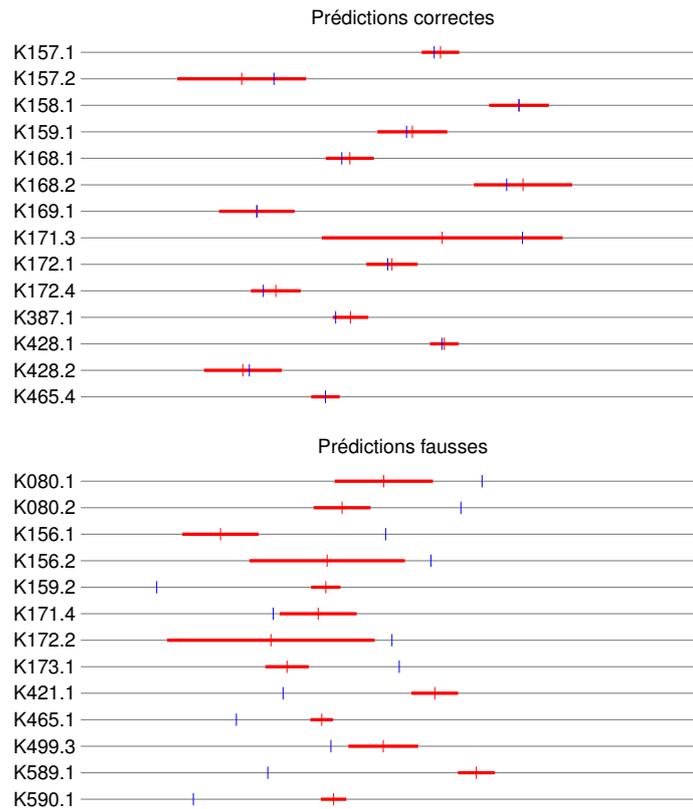


FIGURE 6.6 – Mise en relation des MC de l'annotation de référence (marquées par un trait vertical rouge et entourée par un segment rouge qui correspond à la tolérance de plus ou moins 4 temps) et des MC prédites par le modèle (trait vertical bleu).

premier mouvement du quatuor K157, le pic le plus haut coïncide avec la position de la césure médiane. Le deuxième pic le plus haut, sur le temps 76, est aussi intéressant car il correspond à la demi-cadence qui clôt la section P. Sur le premier mouvement de K168, il n'y a qu'un unique pic qui correspond également à la position de la MC. Sur les deux autres œuvres, le modèle n'identifie pas correctement la MC. Le cas de la prédiction sur le premier mouvement de K421 (voir figure 2.10) est très intéressant. Une MC est prédite à tort sur la mesure 14, soit 40 temps environ avant la MC identifiée dans l'annotation de référence. Cependant, Mozart semble en effet préparer une césure médiane aux mesures précédant la mesure 14. Au tout dernier moment, au lieu d'articuler la césure médiane comme attendu, Mozart décide de moduler dans la tonalité relative (soit *Fa Majeur*) et de continuer un peu plus longtemps sa transition vers S. On peut parler de refus de césure médiane⁶. Sur le premier mouvement de K589 par contre, il n'y a pas réellement de candidat qui se dégage, toutes les probabilités étant plutôt faibles.

6.6 Bilan

De multiples indicateurs musicaux contribuent à l'impression cadentielle de "respiration dans la musique". Les études présentées dans ce chapitre évaluent la pertinence de mettre en

6. Un peu à la manière d'une cadence rompue...

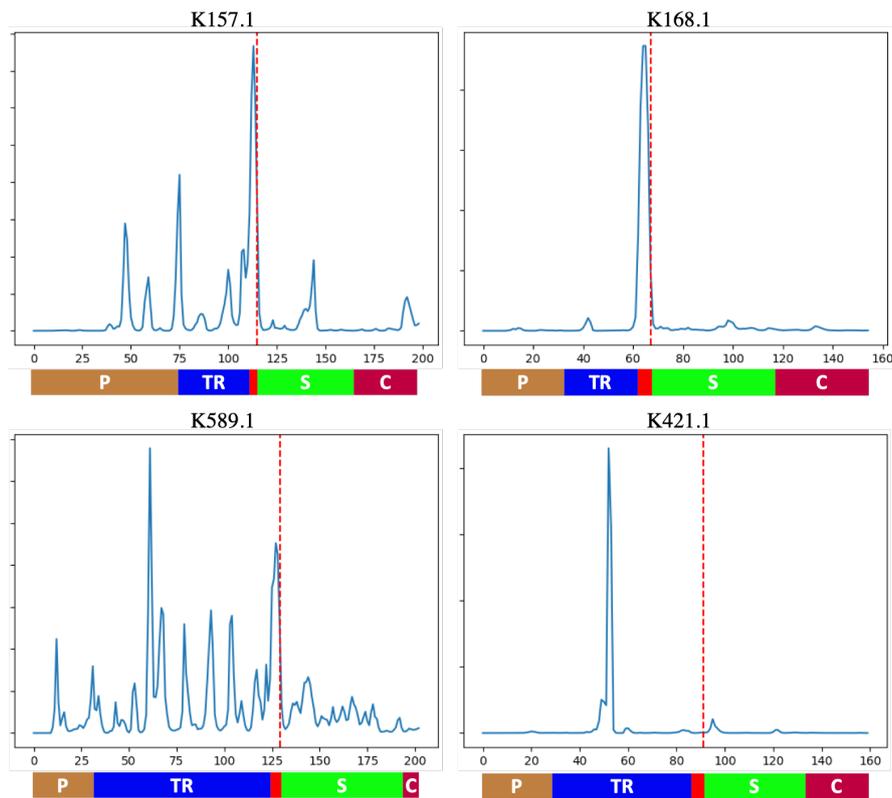


FIGURE 6.7 – Courbes de la probabilité d’être en présence du MC en chaque temps de la partition pour quatre oeuvres du corpus. Les deux courbes du haut correspondent à des prédictions correctes, les deux du bas à des mauvaises prédictions. On recopie en dessous de chaque courbes l’annotation de référence de la structure de la pièce (P/TR/MC/S/C), la ligne verticale rouge pointillée correspond à la position de la MC.

place des descripteurs “haut niveau” pour les cadences, scrutant à la fois le point d’arrivée cadentiel mais aussi son passé proche, à la recherche de sa préparation. Sans recours à une segmentation d’accord, qui est en elle-même un sujet de recherche complexe, ces descripteurs modélisent l’harmonie sous-jacente, la conduite de voix ainsi que d’autres aspects structurels de la musique.

Comme nous l’avons expliqué tout au long de ce chapitre, l’élaboration de ces descripteurs est guidée par la théorie musicale établie sur ce qui se passe à proximité d’une cadence. Ils permettent d’apprendre à un modèle d’apprentissage machine à prédire où sont les cadences, en particulier les PAC, après avoir été entraîné sur des corpus annotés. L’utilisation de ces descripteurs ne saurait se limiter à la détection de cadences et pourrait être adapté à l’automatisation d’autres tâches d’analyse musicale comme par exemple à l’estimation de la forme d’une œuvre ou la texture.

Pour aller plus loin dans la détection des PAC et HC, nous pourrions ajouter d’autres ensembles de descripteurs. Dans ce sens, on pourrait généraliser les descripteurs de type $Z\text{-}\beta\text{-comes-from-}\alpha$ et $Z\text{-}\beta\text{-moves-to-}\alpha$. Peut-être que certains s’avéreraient significatifs. Il serait tout aussi pertinent de trouver des descripteurs plus fiables pour l’analyse harmonique et aussi pour la tonalité, ce qui améliorerait sans doute la performance du modèle. Enfin certains descripteurs pourraient être mis en place afin de mieux caractériser la fin de phrase, notamment

concernant la texture (ou plutôt son changement avant et après la fin de phrase). La méthode utilisée pour estimer les accords Y et X pourrait être comparée à d'autres heuristiques. Ces accords pourraient être considérés comme des périodes de temps continu plutôt que comme des points dans l'espace temporel musical, renforçant ainsi la pertinence harmonique du modèle et probablement la performance de ce dernier sur la prédiction des demi-cadences.

Les résultats de la détection des MC sont encourageants surtout au vu de la taille plus que modeste du corpus et la diversité de réalisation des césures médianes. L'analyse détaillée de la sortie du modèle permet d'identifier d'autres points importants dans l'œuvre et en ce sens peut aider à l'analyse structurelle de cette dernière quand bien même elle échouerait à prédire correctement la MC. Il aurait pu être intéressant d'ajouter à l'ensemble des descripteurs ceux déjà utilisés pour la détection de cadence et présentés à la section 6.2. Nous ne les avons pas ajoutés de prime abord car ceux-ci sont significatifs pour un type de cadence en particulier, or on ne sait pas vraiment à quel type de cadence s'attendre pour une césure médiane.

Pour les deux études, une question est de savoir si l'entraînement d'un réseau de neurones profond sur une description bas niveau de la partition (les notes) en viendrait à reconstituer une extraction de descripteurs similaires à ceux que nous avons choisis dans ce travail – et naturellement comparer les résultats de ces deux approches.

Nous souhaitons également ajouter des descripteurs sur les nuances qui semblent assez bien modéliser l'intensité d'une œuvre (et introduire une notion d'*énergie musicale*) cependant celles-ci ne sont pas présentes dans les fichiers de notre corpus et sont très peu présentes en règle générale.

Dans ce chapitre, nous avons opéré une transformation de la représentation de la partition, en passant d'un niveau très concret, proche du texte musical (les notes à proprement parler) à un niveau moins observable par la modélisation de la partition par une série de descripteurs plus haut niveau. Ce changement de paradigme a aidé à l'analyse musicale automatisée et dans notre cas, à la détection automatisée des cadences et des césures médianes. Ces résultats pourraient eux aussi constituer une nouvelle strate de modélisation de la partition pour permettre une analyse d'une métadonnée encore plus abstraite, l'analyse de la forme musicale et en particulier la forme sonate, ce que nous proposons en guise de conclusion à cette thèse.

Chapitre 7

Conclusion

Tout au long de ce manuscrit, nous avons proposé des modélisations de *concepts musicaux haut-niveau* afin de concevoir des algorithmes et des modèles pour l'analyse musicale computationnelle. Nous avons également constitué trois corpus de partitions symboliques, tout en réfléchissant à leur annotation (chapitre 4). Ces corpus ont servi à évaluer nos modèles et pourront servir pour d'autres études en analyse musicale computationnelle.

Deux approches originales pour l'analyse musicale automatisée ont été explorées pour résoudre deux problèmes distincts.

Une première étude (chapitre 5) s'est penchée sur *l'estimation des tonalités locales* pour parvenir à une prédiction fine des *points de modulations*. Là où les approches antérieures ne s'intéressent essentiellement qu'à l'organisation des notes dans une fenêtre restreinte, notamment par corrélation et pondération des notes, notre proposition essaie de modéliser plus globalement le *système tonal*. En effet, bien que les hauteurs de notes soient une donnée essentielle lorsqu'il s'agit de déterminer la tonalité, ce que nos données confirment, l'ajout d'informations sur la distance harmonique d'une progression harmonique de type dominante vers tonique et surtout l'ajout de pénalités pour rompre l'état d'équilibre stable suggéré par la tonalité permettent d'améliorer la compréhension du problème et les performances pour l'estimation des tonalités locales. Sur des œuvres entières (en l'occurrence les quatuors à cordes de Mozart), la performance est satisfaisante et peut être améliorée en renforçant la modélisation des progressions harmoniques. En revanche, sur des extraits très courts comme ceux du corpus spécialisé, le modèle n'est pas du tout adapté et la performance est mauvaise. Ce résultat particulier peut signifier que plus qu'une approche pour estimer la tonalité locale, nous avons conçu un modèle approché du système tonal en ce sens qu'il propose une valeur de la tension mesurée si on considère le contexte d'une tonalité à un temps précis, contrairement aux algorithmes usant des profils tonaux plus intéressés à estimer la *centricité* d'une tonalité.

Une deuxième étude (chapitre 6) s'est intéressée à la *détection des cadences*. L'approche proposée envisage cette détection comme un problème de *classification des temps musicaux* dans la partition selon leur capacité à être un point d'arrivée cadentiel. L'enjeu principal de cette tâche était de trouver un moyen de modéliser les fins de phrases musicales. Pour ce faire, nous avons mis en place des *descripteurs musicaux* qui nous semblaient caractéristiques de la cadence. Ces techniques améliorent la détection de cadence parfaite, surtout en produisant moins de faux positifs. Les résultats sont plus mitigés pour la détection de demi-cadence, ce qui laisse à penser que nous n'avons pas tout à fait réussi à modéliser cette notion de fin de phrase musicale.

En effet, les descripteurs les plus significatifs portent sur les progressions harmoniques qui sont fixes pour la cadence parfaite. Comme ce n'est pas le cas pour la détection de demi-cadences, les résultats sont moins bons. Malgré tout, nous pensons que l'abstraction proposée par la modélisation et l'extraction de descripteurs haut-niveau, plutôt que de rester au niveau de la note, permet de mieux comprendre les phénomènes qui peuvent aider à déterminer une fin de phrase sur les partitions musicales.

Cette approche a servi de base pour la détection des *césures médianes* dans les expositions de mouvements de forme sonate. La mise en place de descripteurs spécifiques à la césure médiane calqués sur certains signaux issus d'ouvrages de théorie musicale a permis de détecter un peu plus de la moitié des césures médianes du corpus, ce qui est certes peu mais encourageant au vu de la variété des césures médianes dans le corpus et de leur préparation. Cette étape était importante car nous nous confrontions pour la première fois à un des objectifs sous-jacents dans lequel s'inscrit la thèse, la *détection de la forme sonate*. Les résultats sur la détection de la césure médiane sont meilleurs qu'un article d'équipe consacré à la forme sonate [7] : on a donc déjà amélioré la détection de la forme sonate. L'adaptation des travaux effectués dans cette thèse à l'estimation de la forme musicale et en particulier de la forme sonate est clairement la suite logique de cette thèse et pourra donner lieu à de futures recherches. Plus généralement, les résultats satisfaisants sur l'estimation de points de modulation peut sérieusement aider à l'amélioration des résultats sur la détection de la forme sonate. En effet, le parcours tonal est ce qui définit la dialectique de la forme sonate. La détection des cadences et plus généralement des fins de phrases permet de découper la partition en différentes parties qui pourraient être rapprochées des sections de la forme sonate.

Plus généralement, les outils développés dans le cadre de cette thèse pourraient servir tels quels à l'aide à *l'analyse musicale*, même si ces outils manquent encore un peu de fiabilité. Dans le cadre d'une *analyse semi-automatisée*, ils peuvent au moins accélérer le travail de l'analyste qui pourrait se concentrer sur des choses plus compliquées à modéliser tels que le style d'une oeuvre, l'intention du compositeur ou la recherche d'une dialectique sous-jacente. L'avantage de ce genre de technique d'automatisation est que l'algorithme peut sortir des résultats sur de larges corpus de façon très rapide là où il faudrait plusieurs semaines aux musicologues même expérimentés pour cette même tâche. L'extraction de ce genre de descripteurs pourrait aider à observer l'évolution stylistique d'un compositeur ou encore à déterminer ce qui caractérise concrètement un style français baroque autrement que par le fait que "l'harmonie en est brute, sans expression et sentant uniquement son remplissage d'écolier." [116]. Il pourrait aussi être intéressant d'ajouter comme index de certaines oeuvres le parcours tonal, pour pouvoir faire une recherche des partitions par nombre de tonalités empruntées, par parcours tonal relatif ou ce genre de choses. Le choix des tonalités et des modulations n'est pas anodin pour les compositeurs et encore plus pour les oeuvres programmatiques où le choix des tonalités peut avoir une grande signification.

Les modèles présentés dans cette thèse ont été spécifiquement construits pour l'analyse de *musique tonale de la période classique*. Nous pensons qu'ils peuvent être adaptés sans trop de modifications pour la musique "savante" tonale d'autres périodes, comme la musique baroque ou la musique romantique. La détection de cadence sur les fugues de Bach a obtenu des bons résultats bien que pourtant j'imaginai que le caractère contrapuntique de cette musique pouvait rendre difficile la détection des fins de phrases. On a là un cas concret d'aide à la segmentation de la partition en phrases malgré l'avancée inexorable de cette musique. Pour la musique romantique et post-romantique allemande, le défi serait d'adapter le modèle d'estimation des

tonalités locales. En effet, on peut y trouver beaucoup de zones tonales instables et d'inserts de modalités qui affaiblissent le signal d'ancrage dans la tonalité. De plus, les modulations ne se font plus seulement dans des tonalités adjacentes, mais peuvent aussi se réaliser chromatiquement ou dans des tonalités enharmoniques qui, si l'on se réfère à la table de Weber, ont un coût vraiment très élevé. Il faudrait trouver un autre système de mise en relation des tonalités.

Il semble plus incongru d'utiliser ces modèles pour les *musiques modales et atonales*. En effet, ces deux modèles participent modestement à la modélisation du système tonal tel qu'il a été décrit dans la section 2.1, et bien que certaines des caractéristiques évoquées puissent se retrouver partiellement dans la musique non tonale, l'analyse musicale pour ces types de musique doit être pensée différemment. Cependant, nous pensons que l'approche utilisée pour la détection de cadences peut être adaptée pour la recherche de fins de phrases (plus que les cadences, si jamais elle existent pour ces répertoires) pour tout type de musique. Les descripteurs choisis devront être forcément adaptés en conséquence.

Il est plus compliqué d'adapter l'algorithme d'estimation des points de modulation, en particulier pour les *musiques contemporaines*, telles que musiques atonales ou sérielles. En effet, les progressions harmoniques stabilisatrices de type dominante vers tonique ne sont pas aussi souvent utilisées, quand ces fonctions existent ce qui est loin d'être évident pour beaucoup des œuvres de ces répertoires. De plus, les relations entre les tonalités ne sont probablement pas les mêmes. Et est-il tout simplement pertinent de parler d'estimation de tonalité pour de la musique non-tonale? Probablement pas. Concernant les musiques contemporaines populaires, la situation est plus complexe : selon Tymoczko [132], jazz, pop et autres musiques appartiennent à une "période tonale étendue" (*extended common practice period*). Je pense à titre personnel que les outils d'analyse musicale "traditionnels" ne sont pas toujours pertinents pour ce style de musique. La détection de points de modulation me semble en particulier futile tant le phénomène de modulation y est rare. Pour la détection de cadence, cela peut être différent. Dans des structures en boucle de 4, 8 ou 16 mesures sans vraiment de coupure entre les sections, on peut certes réussir à identifier des fins de phrases. Par contre, je doute que mon modèle y soit efficace. En effet, les musiques contemporaines populaires sont à mon sens plus modales que représentatives du système tonal, ainsi la fin de phrase ne s'achèvera pas forcément par une progression harmonique $V \rightarrow I$.

Mais encore une fois, nous pensons que la méthode utilisée peut-être adaptée à d'autres problèmes musicaux qui peuvent être modélisés par des mesures de distance ou de coût. Certains programmes de génération automatisée de musique pourraient assigner à chaque note possible en un temps donné un coût en fonction du contexte dans lequel se trouve la dite note. La ligne mélodique choisie serait alors celle qui minimise les coûts. Pour une méthode de ce type, nul doute que l'analyse musicale automatique est d'une grande aide. Nos outils peuvent aider à la construction d'une structure adéquate et d'un parcours tonal cohérent.

Bibliographie

- [1] Bret J. Aarden. *Dynamic Melodic Expectancy*. PhD thesis, The Ohio State University, 2003.
- [2] Claude Abromont and Eugène De Montalembert. *Guide de la théorie de la musique*. Fayard Henry Lemoine, Paris, 2010.
- [3] Christian Accaoui. *Eléments d'esthétique musicale : notions, formes et styles en musique*. Actes sud, 2011.
- [4] Christian Accaoui and Maxime Joos. *Tonalité*. In [3], 2011.
- [5] Joshua Albrecht and Daniel Shanahan. The use of large corpora to train a new type of key-finding algorithm : An improved treatment of the minor mode. *Music Perception : An Interdisciplinary Journal*, 31(1) :59–67, 2013.
- [6] Edward Aldwell, Carl Schachter, and Allen Cadwallader. *Harmony and Voice Leading*. Cengage Learning, 2019.
- [7] Pierre Allegraud, Louis Bigo, Laurent Feisthauer, Mathieu Giraud, Richard Groult, Emmanuel Leguy, and Florence Levé. Learning Sonata Form Structure on Mozart's String Quartets. *Transactions of the International Society for Music Information Retrieval*, 2(1) :82–96, 2019.
- [8] Moreno Andreatta, Gérard Assayag, Myriam Desainte-Catherine, Jean-Louis Giavitto, Mathieu Giraud, Yann Orlarey, and Anne Sèdes. L'informatique musicale : informatique et musique en interdisciplinarité. In *Journées d'Informatique Musicale (JIM 2019)*, 2019.
- [9] Claire Arthur. Taking harmony into account : The effect of harmony on melodic probability. *Music Perception*, 34(4) :405–423, 2017.
- [10] Antoine Auda. Le tactus principe générateur de l'interprétation de la musique polyphonique classique. 1950.
- [11] Richard Bellman. *Dynamic Programming*. Dover Publications, 1957.
- [12] Héctor Bellmann. About the determination of key of a musical excerpt. In *International Symposium on Computer Music Modeling and Retrieval (CMMR 2006)*, pages 76–91, 2006.
- [13] Ian Bent and William Drabkin. *Analysis*. W.W. Norton & Company, 1987.
- [14] Louis Bigo, Laurent Feisthauer, Mathieu Giraud, and Florence Levé. Relevance of Musical Features for Cadence Detection. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*, 2018.
- [15] Louis Bigo, Mathieu Giraud, Richard Groult, Nicolas Guiomard-Kagan, and Florence Levé. Sketching Sonata Form Structure in Selected Classical String Quartets. In *International Society for Music Information Retrieval Conference (ISMIR 2017)*, pages 752–759, 2017.
- [16] Ann Blombach. Phrase and cadence : A study of terminology and definition. *Journal of Music Theory Pedagogy*, 1 :225–51, 1987.

- [17] William E. Caplin. *Classical Form : A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford University Press, 1998.
- [18] William E. Caplin. The classical sonata exposition : Cadential goals and form-functional plans. *Tijdschrift voor Muziektheorie*, 6(3) :195–209, 2001.
- [19] William E. Caplin. The classical cadence : Conceptions and misconceptions. *Journal of the American Musicological Society*, 57 :51–117, 2004.
- [20] William E. Caplin, James Hepokoski, and James Webster. *Musical Form, Forms & Formenlehre – Three Methodological Reflections*. Pieter Bergé, Leuven University Press, 2009.
- [21] Benoit Catteau, Jean-Pierre Martens, and Marc Leman. A probabilistic framework for audio-based tonal key and chord recognition. In Reinhold Decker and Hans J. Lenz, editors, *Advances in Data Analysis*, pages 637–644, Springer, Berlin, Heidelberg, 2007.
- [22] Wei Chai and Barry Vercoe. Detection of key change in classical piano music. In *International Conference on Music Information Retrieval (ISMIR 2005)*, pages 468–473, 2005.
- [23] Tsung-Ping Chen and Li Su. Functional harmony recognition of symbolic music data with multi-task recurrent neural networks. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*, pages 90–97, 2018.
- [24] Tsung-Ping Chen and Li Su. Harmony transformer : Incorporating chord segmentation into harmony recognition. In *International Society for Music Information Retrieval Conference (ISMIR 2019)*, pages 259–267, 2019.
- [25] Elaine Chew. *Towards a Mathematical Model of Tonality*. PhD thesis, MIT, Cambridge, MA, 2000.
- [26] Elaine Chew. The spiral array : An algorithm for determining key boundaries. In *Music and Artificial Intelligence*, pages 18–31. Springer, 2002.
- [27] François Chollet. *Deep learning with Python*. Manning Publications Company, 2017.
- [28] François Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015.
- [29] Alexandre Choron and François-Joseph-Marie Fayolle. *Dictionnaire historique des musiciens artistes et amateurs, morts ou vivans*. Valade ou Lenormant, Paris, 1810.
- [30] Tom Collins, Andreas Arzt, Sebastian Flossmann, and Gerhard Widmer. SIARCT-CFP : Improving precision and the discovery of inexact musical patterns in point-set representations. In *International Society for Music Information Retrieval Conference (ISMIR 2013)*, pages 549–554, 2013.
- [31] Nathaniel Condit-Schultz, Yaolong Ju, and Ichiro Fujinaga. A flexible approach to automated harmonic analysis : Multiple annotations of chorales by Bach and Prætorius. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*, pages 66–73, 2018.
- [32] Darrell Conklin and Christina Anagnostopoulou. Representation and discovery of multiple viewpoint patterns. In *International Computer Music Conference (ICMC 2001)*, pages 479–485, 2001.
- [33] Darrell Conklin and Ian H Witten. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1) :51–73, 1995.
- [34] Julie E Cumming, Cory McKay, Jonathan Stuchbery, and Ichiro Fujinaga. Methodologies for creating symbolic corpora of Western music before 1600. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*, pages 491–498, 2018.

- [35] Michael Scott Cuthbert and Christopher Ariza. music21 : A toolkit for computer-aided musicology and symbolic music data. In *International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 637–642, 2010.
- [36] Carl Czerny. *School of Practical Composition*. R. Cocks, London, 1848.
- [37] Carl Dahlhaus. *La tonalité harmonique : étude des origines*. Collection Musique, musicologie. Mardaga, 1993.
- [38] W. Bas De Haas, Martin Rohrmeier, Remco C. Veltkamp, and Frans Wiering. Modeling harmonic similarity using a generative grammar of tonal harmony. In *International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 549–554, 2009.
- [39] Jacob deGroot Maggetti, Timothy de Reuse, Laurent Feisthauer, Samuel Howles, Yaolong Ju, Suzaka Kokubu, Sylvain Margot, Néstor Nápoles López, and Finn Upham. Data quality matters : Iterative corrections on a corpus of Mendelssohn string quartets and implications for MIR analysis. In *International Society for Music Information Retrieval Conference (ISMIR 2020)*, 2020.
- [40] Nikolai Diletsky. *Idea grammatikii musikiyskoy*. Moskow, 1679.
- [41] J. Stephen Downie, Kris West, Andreas F. Ehmann, and Emmanuel Vincent. The 2005 music information retrieval evaluation exchange (MIREX 2005) : Preliminary overview. In *International Conference on Music Information Retrieval (ISMIR 2005)*, pages 320–323, 2005.
- [42] Ben Duane. Melodic patterns and tonal cadences : Bayesian learning of cadential categories from contrapuntal information. *Journal of New Music Research*, 48(3) :197–216, 2019.
- [43] Ben Duane and Joseph Jakubowski. Harmonic clusters and tonal cadences : Bayesian learning without chord identification. *Journal of New Music Research*, 47(2) :143–165, 2018.
- [44] Fabian Pedregosa et al. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.
- [45] Laurent Feisthauer, Louis Bigo, and Mathieu Giraud. Modeling and learning structural breaks in sonata forms. In *International Society for Music Information Retrieval Conference (ISMIR 2019)*, 2019.
- [46] Laurent Feisthauer, Louis Bigo, Mathieu Giraud, and Florence Levé. Vers une recherche automatisée du parcours tonal. *Les sciences de la musique*, 2019.
- [47] Laurent Feisthauer, Louis Bigo, Mathieu Giraud, and Florence Levé. Estimating keys and modulations in musical pieces. In *Sound and Music Computing Conference (SMC 2020)*, Torino, Italy, 2020.
- [48] François-Joseph Fétis. *Traité complet de la théorie et de la pratique de l’harmonie*. Paris : Schlesinger, 1844.
- [49] Marius Flothuis. *Mozarts Streichquartette : Ein musikalischer Werkführer*. C. H. Beck, 1998.
- [50] Allen Forte. *The Structure of Atonal Music*. Yale University Press, 1973.
- [51] Francesco Foscari, Florent Jacquemard, and Raphaël Fournier-S’niehotta. A diff procedure for music score files. In *Digital Libraries for Musicology (DLfM 2019)*, pages 58–64, 2019.
- [52] Francesco Foscari, Florent Jacquemard, and Philippe Rigaux. Modeling and learning rhythm structure. In *Sound and Music Computing Conference (SMC 2019)*, pages 566–572, 2019.
- [53] Zoltán Gárdonyi and Hubert Nordhoff. *Harmonik*. Mösel, 1990.

- [54] Mathieu Giraud, Richard Groult, and Emmanuel Leguy. Dezzrann, a web framework to share music analysis. In *International Conference on Technologies for Music Notation and Representation (TENOR 2018)*, pages 104–110, 2018.
- [55] Mathieu Giraud, Richard Groult, Emmanuel Leguy, and Florence Levé. Computational fugue analysis. *Computer Music Journal*, 39(2), 2015.
- [56] Mathieu Giraud, Richard Groult, and Florence Levé. Subject and counter-subject detection for analysis of the Well-Tempered Clavier fugues. In *International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)*, pages 661–673, 2012.
- [57] Robert O. Gjerdingen. *Music in the Galant Style*. Oxford University Press, 2007.
- [58] Yoel Greenberg. Of beginnings and ends : A corpus-based inquiry into the rise of the recapitulation. *Journal of Music Theory*, 61(2) :171–200, 2017.
- [59] Nicolas Guiomard-Kagan, Mathieu Giraud, Richard Groult, and Florence Levé. Comparing voice and stream segmentation algorithms. In *International Society for Music Information Retrieval Conference (ISMIR 2015)*, pages 493–499, 2015.
- [60] Nicolas Guiomard-Kagan, Mathieu Giraud, Richard Groult, and Florence Levé. Improving voice separation by better connecting contigs. In *International Society for Music Information Retrieval Conference (ISMIR 2016)*, pages 192–198, 2016.
- [61] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. *Implementing Methods for Analysing Music Based on Lerdahl and Jackendoff’s Generative Theory of Tonal Music*, pages 221–249. In Meredith [85], 2016.
- [62] Johann David Heinichen. *Der General-Bass in der Composition*. Dresden, 1711.
- [63] James Hepokoski. Beyond the sonata principle. *Journal of the American Musicological Society*, 55(2) :91, 2002.
- [64] James Hepokoski and Warren Darcy. The medial caesura and its role in the eighteenth-century sonata exposition. *Music Theory Spectrum*, 19(2) :115–154, 1997.
- [65] James Hepokoski and Warren Darcy. *Elements of Sonata Theory : Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. Oxford University Press, 2006.
- [66] Lejaren Arthur Hiller and Leonard Maxwell Isaacson. *Experimental music : composition with an electronic computer*. McGraw-Hill, 1959.
- [67] David Huron. Music information processing using the Humdrum toolkit : Concepts, examples, and lessons. *Computer Music Journal*, 26(2) :11–26, 2002.
- [68] Özgür Izmirli. Localized key finding from audio using nonnegative matrix factorization for segmentation. In *International Conference on Music Information Retrieval (ISMIR 2007)*, pages 195–200, 2007.
- [69] Vladimir Jankélévitch. *La Musique et l’ineffable*. Éd. du Seuil, Paris.
- [70] Yaolong Ju, Samuel Howes, Cory McKay, Nathaniel Condit-Schultz, Jorge Calvo-Zaragoza, and Ichiro Fujinaga. An interactive workflow for generating chord labels for homorhythmic music in symbolic formats. In *International Society for Music Information Retrieval Conference (ISMIR 2019)*, pages 862–869, 2019.
- [71] Alec Hyatt King. *La Musique de chambre de Mozart*. Arles : Actes Sud, 1968.
- [72] Hendrik Vincent Koops, Bas de Haas, John Ashley Burgoyne, Jeroen Bransen, Anna Kent-Muller, and Anja Volk. Annotator subjectivity in harmony annotations of popular music. *Journal of New Music Research*, 48(3) :232–252, 2019.

- [73] Stefan Kostka and Dorothy Payne. *Tonal Harmony*. McGraw-Hill Education, Boston, 2008.
- [74] Carol L. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford Psychology Series. Oxford University Press, USA, 1990.
- [75] Carol L. Krumhansl and Edward J. Kessler. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89(4) :334–368, 1982.
- [76] Carol L. Krumhansl and Roger N. Shepard. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology : Human Perception and Performance*, 5(4) :579–594, 1979.
- [77] Steve Larson. Recapitulation recomposition in the sonata-form first movements of Haydn’s string quartets : Style change and compositional technique. *Music Analysis*, 22(1-2) :139–177, 2003.
- [78] Fred Lerdahl. Tonal pitch space. *Music Perception*, 5 :315–350, 01 1988.
- [79] Fred Lerdahl and Ray Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1983.
- [80] Hugh C. Longuet-Higgins and Mark Steedman. On interpreting Bach. *Mental processes : Studies in cognitive science*, pages 221–241, 1971.
- [81] Ada Lovelace. A sketch of the analytical engine, with notes by the translator. *Scientific Memoirs*, 3 :666–731, 1843.
- [82] Nathan John Martin and Julie Pedneault-Deslauriers. *The Mozartean Half Cadence*, pages 185–213. In Neuwirth and Bergé [96], 2015.
- [83] Adolf Bernhard Marx. *Die Lehre von der musikalischen Komposition, praktisch theoretisch*. Breitkopf und Härtel, 1837-1847.
- [84] Lesley Mearns, Emmanouil Benetos, and Simon Dixon. Automatically detecting key modulations in J.-S. Bach chorale recordings. In *Sound and Music Computing Conference (SMC 2011)*, pages 25–32, 2011.
- [85] David Meredith. *Computational Music Analysis*. Springer, 2015.
- [86] Gianluca Micchi, Mark Gotham, and Mathieu Giraud. Not all roads lead to Rome : Pitch representation and model architecture for automatic harmonic analysis. *Transactions of the International Society for Music Information Retrieval*, 3(1) :42–54, 2020.
- [87] Sylvette Milliot. *Le Quatuor (Que sais-je ?)*. Presses universitaires de France, 1986.
- [88] Jan Miyake. *The Role of Multiple New-key Themes in Selected Sonata-form Exposition*. PhD thesis, Univ. of New York, 2004.
- [89] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.
- [90] Fabian C. Moss, Markus Neuwirth, Daniel Harasim, and Martin Rohrmeier. Statistical characteristics of tonal harmony : A corpus study of Beethoven’s string quartets. *PLOS ONE*, 14(6) :1–16, 2019.
- [91] Leopold Mozart. Korrespondenz. <http://data.onb.ac.at/dtl/5771872>, 1785.
- [92] Néstor Nápoles López, Claire Arthur, and Ichiro Fujinaga. Key-finding based on a hidden Markov model and key profiles. In *Digital Libraries for Musicology (DLfM 2019)*, pages 33–37. ACM, 2019.
- [93] Néstor Nápoles López, Laurent Feisthauer, Florence Levé, and Ichiro Fujinaga. On local keys, modulations, and tonicizations : A dataset and methodology for evaluating changes of key. In *Digital Libraries for Musicology (DLfM 2020)*, pages 18–26. ACM, 2020.

- [94] Néstor Nápoles López and Ichiro Fujinaga. Harmalysis : A language for the annotation of roman numerals in symbolic music representations. In *Music Encoding Conference (MEC 2020)*, pages 83–85, 2020.
- [95] Eugene Narmour. *The Analysis and Cognition of Melodic Complexity : The Implication-Realization Model*. University of Chicago Press, 1992.
- [96] Markus Neuwirth and Pieter Bergé, editors. *What Is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*. Leuven University Press, 2015.
- [97] Markus Neuwirth, Daniel Harasim, Fabian C. Moss, and Martin Rohrmeier. The Annotated Beethoven Corpus (ABC) : A dataset of harmonic analyses of all Beethoven string quartets. *Frontiers in Digital Humanities*, 5 :16, 2018.
- [98] Hélène Papadopoulos and Geoffroy Peeters. Local key estimation based on harmonic and metric structures. In *International Conference on Digital Audio Effects (DAFx-09)*, pages 408–415, 2009.
- [99] Johan Pauwels and Jean-Pierre Martens. Combining musicological knowledge about chords and keys in a simultaneous chord and local key estimation system. *Journal of New Music Research*, 43(3) :318–330, July 2014.
- [100] Steffen Pauws. Musical key extraction from audio. In *International Conference on Music Information Retrieval (ISMIR 2004)*.
- [101] Marcus Thomas Pearce. *The construction and evaluation of statistical models of melodic structure in music perception and composition*. PhD thesis, City University London, 2005.
- [102] Alain Poirier. *Guide de la Musique de Chambre - dir. François-René Tranchefort*. Fayard, 1989.
- [103] Hendrik Purwins, Benjamin Blankertz, and Klaus Obermayer. A new method for tracking modulations in tonal music in audio data format. In *International Joint Conference on Neural Networks (IJCNN 2000)*, volume 6, pages 270–275, 2000.
- [104] Ian Quinn and Christopher W. White. Corpus-derived key profiles are not transpositionally equivalent. *Music Perception*, 34(5) :531–540, 2017.
- [105] Jean-Philippe Rameau. *Traité de l'harmonie réduite à ses principes naturels*. J. B. C. Ballard, Paris, 1722.
- [106] Leonard Ratner. *Classical Music : Expression, Form, and Style*. Schirmer, 1980.
- [107] Max Reger. *Beiträge zur Modulationslehre*. Kahnt, 1904.
- [108] Anton Reicha. *Traité de haute composition musicale*. A. Diabelli, 1824.
- [109] Mark Richards. Sonata form and the problem of second-theme beginnings. *Music Analysis*, 32(1) :3–45, 2013.
- [110] Marc Rigaudière. *La théorie musicale germanique du XIX^e siècle et l'idée de cohérence*. Société Française de Musicologie, 2009.
- [111] Nikolay Rimski-Korsakov. *Practical Manual of Harmony*. A. Büttner, 1886.
- [112] Thomas Rocher, Matthias Robine, Pierre Hanna, and Laurent Oudre. Concurrent estimation of chords and keys from audio. In *International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 141–146, 2010.
- [113] Martin Rohrmeier. A generative grammar approach to diatonic harmonic structure. In *Sound and Music Computing Conference (SMC 2007)*, pages 97–100, 2007.
- [114] Martin Rohrmeier and Markus Neuwirth. *Towards a Syntax of the Classical Cadence*, pages 287–338. In Neuwirth and Bergé [96], 2015.

- [115] Charles Rosen. *Sonata Forms*. W. W. Norton, 1980.
- [116] Jean-Jacques Rousseau. *Lettre sur la musique française*. 1753.
- [117] Craig Stuart Sapp. Visual hierarchical key analysis. *Computers in Entertainment*, 3(4) :1–19, 2005.
- [118] Craig Stuart Sapp. *Computational Methods for the Analysis of Musical Structure*. PhD thesis, Stanford University, 2011.
- [119] Heinrich Schenker. *Harmonielehre*. Universal Edition, 1906.
- [120] Heinrich Schenker. *Der freie Satz*. Universal Edition, 1935.
- [121] Janet Schmalfeldt. Cadential processes : The evaded cadence and the “one more time” technique. *Journal of Musicological Research*, 12(1-2) :1–52, 1992.
- [122] David R. W. Sears, William E. Caplin, and Stephen McAdams. Perceiving the classical cadence. *Music Perception*, 31(5) :397–417, 2014.
- [123] David R. W. Sears, Marcus T. Pearce, William E. Caplin, and Stephen McAdams. Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *Journal of New Music Research*, 47(1) :29–52, 2017.
- [124] David R. W. Sears, Jacob Spitzer, William E. Caplin, and Stephen McAdams. Expecting the end : Continuous expectancy ratings for tonal cadences. *Psychology of Music*, 48(3) :358–375, 2018.
- [125] Anna Selway, Hendrik Vincent Koops, Anja Volk, David Bretherton, Nicholas Gibbins, and Richard Polfreman. Explaining harmonic inter-annotator disagreement using Hugo Riemann’s theory of ‘harmonic function’. *Journal of New Music Research*, 49(2) :136–150, 2020.
- [126] Eline A. Smit, Felix A. Dobrowohl, Nora K. Schaal, Andrew J. Milne, and Steffen A. Herff. Perceived emotions of harmonic cadences. *Music & Science*, 3 :2059204320938635, 2020.
- [127] Rémy Stricker. *Franz Liszt : Les ténèbres de la gloire*. Gallimard, 1993.
- [128] Peter I. Tchaikovsky. *Guide to the Practical Study of Harmony*. P. Jurgenson, Moscow, 1872.
- [129] David Temperley. What’s key for key ? the Krumphansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, 17(1) :65–100, 1999.
- [130] David Temperley. *Music and probability*. MIT Press, 2007.
- [131] R. Larry Todd. Mendelssohn(-Bartholdy), (Jacob Ludwig) Felix. *Grove Music Online*, 2000.
- [132] Dmitri Tymoczko. *A Geometry of Music : Harmony and Counterpoint in the Extended Common Practice*. Oxford University Press, 2011.
- [133] Peter van Kranenburg and Folgert Karsdorp. Cadence detection in western traditional stanzaic songs using melodic and textual features. In *International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 391–396, 2014.
- [134] Marc Vignal. *Dictionnaire de la musique*. In extenso. Larousse, Paris.
- [135] Piet G. Vos and Erwin W. Van Geenen. A parallel-processing key-finding model. *Music Perception : An Interdisciplinary Journal*, 14(2) :185–223, 1996.
- [136] Katherine Walker. *Er Hat Geschmack’ : Shifting Connotations Of Taste In The Discourse Surrounding W. A. Mozart*. PhD thesis, Cornell University, 2014.
- [137] Gottfried Weber. *Versuch einer geordneten Theorie der Tonsetzkunst*. B. Schotts Söhne, 1817–21.

- [138] Christof Weiß and Julian Habryka. Chroma-based scale matching for audio tonality analysis. In *Conference on Interdisciplinary Musicology (CIM 2014)*, pages 168–173, 2014.
- [139] Christopher W. White and Ian Quinn. The Yale-classical archives corpus. *Empirical Musicology Review*, 11(1), 2016.

Cette thèse s'inscrit dans le domaine de l'informatique musicale et plus particulièrement de l'analyse musicale computationnelle. Ces études ont pour but de générer des annotations musicales plus ou moins haut-niveau sur une partition, en particulier pour comprendre la genèse, le geste compositionnel ou encore sa place dans l'œuvre globale d'un compositeur. Cette thèse propose de nouvelles approches basées sur la modélisation, l'algorithmique et l'apprentissage machine pour modéliser la *tonalité*, un système musical qui permet de hiérarchiser et contextualiser les notes, ainsi que les *cadences*, qui sont les processus de clôture des phrases musicales. Nous souhaitons ainsi aider à l'analyse d'œuvres en forme sonate.

Nous présentons trois corpus établis durant la thèse – quatuors à cordes de Mendelssohn, quatuors à cordes de Mozart et exemples de modulation – et discutons les étapes et problématiques d'un tel travail. Nous concevons un algorithme d'estimation des tonalités en tout point de la partition, utilisant une nouvelle modélisation du système tonal pour identifier les points de modulation. Il estime à chaque temps trois signaux musicaux : l'ancrage dans la tonalité, la compatibilité des notes avec une tonalité donnée et la proximité entre les tonalités. L'algorithme est évalué sur des corpus de Mozart et de modulations. Nous établissons un algorithme de détection des cadences par l'extraction de descripteurs haut-niveau caractéristiques de la présence d'un point d'arrivée cadentiel sur la partition musicale. Nous étudions la significativité de chacun de ces descripteurs, puis ceux-ci servent à entraîner un algorithme d'apprentissage qui classe chaque temps de la partition comme un point d'arrivée cadentiel ou non. Cet algorithme est évalué sur un corpus de fugues de Bach et de quatuors à cordes d'Haydn et est adapté à la détection d'une cadence particulière significative pour la forme sonate, la césure médiane.

Cette thèse contribue donc à la modélisation informatique de concepts musicaux haut-niveau comme la tonalité et la forme musicale.

This PhD thesis is in Music Information Retrieval (MIR). More specifically, Computational Music Analysis (CMA) studies aim to generate musical annotations on scores to understand how it has been composed and how it characterizes the style of a composer. In this dissertation, we propose original models and algorithms, inspired by machine learning and music theory, to model tonality, a musical system that organizes pitches, and cadences, which are processes of closing a musical sentence. These models should help to analyze pieces in sonata form.

Three corpora were established during this PhD – Mendelssohn string quartets, Mozart string quartets, and modulation excerpts. We discuss the methodology of corpus creation along with the issues encountered. We design an algorithm for local key detection, aiming to improve previous studies by modeling the tonal system to find modulation points and improve the results of the state of the art. It computes three musical signals on each beat of the score : pitch compatibility, harmonic anchoring, and key relationship proximity. This algorithm is evaluated on the Mozart corpus and the modulation corpus. We also propose an algorithm to detect cadences. It extracts high-level musical features that characterize the presence of a cadential arrival point on the score. We study the significance of the features and use them to train a machine learning algorithm that classifies each beat of the score as a cadential arrival point or not. This algorithm is evaluated on a Bach fugue corpus and Haydn string quartets corpus. It is also adjusted and used to find a peculiar cadence found in sonata forms, the medial caesura.

The thesis thus aims to improve computational models of high-level music concepts such as harmony and musical form.